(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 December 2003 (11.12.2003)

PCT

(10) International Publication Number
WO 03/102191 A1

(51) International Patent Classification[7]: C12N 15/52, A61K 48/00, C12N 15/79, 5/10

(21) International Application Number: PCT/GB03/02338

(22) International Filing Date: 30 May 2003 (30.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0212666.2 31 May 2002 (31.05.2002) GB

(71) Applicant *(for all designated States except US)*: **THE SECRETARY OF STATE FOR DEFENCE** [GB/GB]; DSTL, Porton Down, Salisbury, Wiltshire SP4 0JQ (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PRIOR, Joann, Lisa** [GB/GB]; DSTL, Porton Down, Salisbury, Wiltshire SP4 0JQ (GB). **PRIOR, Richard, Geoffrey** [GB/GB]; DSTL, Porton Down, Salisbury, Wiltshire SP4 0JQ (GB). **HITCHEN, Paul, Gareth** [GB/GB]; DSTL, Porton Down, Salisbury, Wiltshire SP4 0JQ (GB). **DELL, Anne** [GB/GB]; DSTL, Porton Down, Salisbury, Wiltshire SP4 0JQ (GB). **TITBALL, Richard, William** [GB/GB]; DSTL, Porton Down, Salisbury, Wiltshire SP4 0JQ (GB).

(74) Agent: **BECKHAM, Robert, William**; D/IPR Formalities Section, Poplar 2, MOD Abbey Wood #2218, Bristol BS34 8JH (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/102191 A1

(54) Title: IMMUNOGENIC SEQUENCES

(57) Abstract: The application relates to nucleic acids which encode enzymes responsible for the production of the O-antigen of *Francisella tularensis*, and their use as or in the production of vaccines and in diagnosis.

Immunogenic Sequences

The present invention relates to nucleic acid sequences, in particular genes that encode the enzymes which produce the O-antigen of *Francisella tularensis*, and their use as or in the production of vaccines and in diagnosis.

Francisella tularensis is a small Gram-negative coccobacillus, which causes the zoonotic disease Tularemia. According to Bergey's manual of systematic bacteriology the genus *Francisella* contains two species: *F. tularensis* and *Francisella novicida*. However, recently several workers have suggested that *F. novicida* be considered a subspecies of *F. tularensis* (Hollis DG, et al., *J.Clin.Micro.* **27**: 1601-1608). The closely related bacterium *Yersinia philomiragia* is now also considered a member of the genus *Francisella*, due to its high degree of relatedness at the DNA level. There are several proposed subspecies of *F. tularensis* other than *novicida*; these are: subspecies *tularensis*, subspecies *holarctica* and subspecies *mediaasiatica*. The subspecies *tularensis* and *holarctica* can be identified on the basis of virulence, citrulline ureidase activity and acid production from glycerol (Olsufjev NG, et al. (1959) *J.Hyg.Epidemiol.Microbiol.Immunol.* **3**: 138-149. *Francisella tularensis* subspecies *mediaasiatica* is predominantly found in central asian republics of the former USSR. Strains of this subspecies possess citrulline ureidase activity, and are able to ferment glycerol, but are less virulent than strains of *F. tularensis* subspecies *tularensis* in rabbit.

Tularemia is a disease occurring in the northern hemisphere; with cases frequently found in Europe, N. America, Asia, N. Russia and Japan. Rodents are thought to be the main reservoir of the bacteria, with ticks as one of the main vectors.

The lipopolysaccharide (LPS) of Gram-negative bacteria is the major component of the outer membrane. The molecule is composed of 3 regions, lipid-A, which is embedded in the outer membrane and has a conserved structure between species, and two

polysaccharides, the core oligosaccharide which can vary in complexity between species, and the O-antigen which is a very polymorphic structure (Kenne L, et al. (1983) Bacterial Polysaccharides *The polysaccharides.* Academic Press, pp. 287-362). The LPS molecule is thought to be required by the bacteria for protection against serum killing (Whitfield C, et al, (1997) *Mol.Micro.* **23**: 629-638) and cationic antimicrobial peptides (Groisman EA. (1994). *Trends.Microbiol.* **2**: 444-449).

The structure and immunogenicity of LPS isolated from the less virulent *F. tularensis* subspecies *holarctica* strains has been studied to some degree (Dreisbach VC, et al. (2000) *Infect.Immun* **68**: 1988-1996). Animals immunised with this LPS are protected against a subspecies *holarctica* strain challenge (Fulop MJ, et al. (1995). *Vaccine* **13**: 1220-1225), but not a subspecies *tularensis* strain challenge (Fulop MJ, et al. (2001). *Vaccine* **19**: 4465-4472). However, the LPS from a subspecies *holarctica* strain appears to be less toxic than other Gram-negative LPS and its O-antigen contains rare sugars which are related in structure to those found in *Pseudomonas aeruginosa* 06 and *Shigella dysenteriae* type 7.

There are no reports of LPS isolation from the more virulent subspecies *tularensis* strains.

When LPS structure is studied in other species, it is frequently observed that the only difference in structure between strains is the composition of the O-antigen. Therefore, it would be useful to elucidate the structure of the O-antigen part of the LPS molecule in virulent subspecies in order to provide the basis for diagnostic tests and also to allow it to be produced recombinantly, to avoid handling a pathogenic organism.

However, the genetic basis of O-antigen expression is complex; in most bacteria the genes required for production of a complete O-antigen are located in a cluster on the bacterial chromosome. Therefore identification and isolation of genes responsible for the O-antigen is not straightforward. Furthermore, the identification and isolation of LPS from

virulent strains is further complicated because it is difficult to stain using conventional methods.

The applicants have now determined the genetic basis of O-antigen production in *F. tularensis* subspecies *tularensis*. Furthermore, they have established the efficacy of LPS from various *F. tularensis* strains as a vaccine.

According to the present invention there is provided a nucleic acid which encodes a series of enzymes or enzyme fragments which, when expressed together in a cell, are able to produce an immunogenic moiety able to produce an immune response in an animal to which it is administered, which response is protective against *Francisella tularensis* infection, said nucleic acid encoding at least some of the enzymes of SEQ ID NOS 3-17, or modifications thereof.

The expression "modification" refers to sequences of amino acids, which differ from the base sequence from which they are derived in that one or more amino acids within the sequence are substituted for other amino acids. Amino acid substitutions may be regarded as "conservative" where an amino acid is replaced with a different amino acid with broadly similar properties. Non-conservative substitutions are where amino acids are replaced with amino acids of a different type. Broadly speaking, fewer non-conservative substitutions will be possible without altering the biological activity of the polypeptide. Suitably modifications will be at least 60% indentical, preferably at least 75% identical, and more preferably at least 90% identical to the base sequence.

Identity in this instance can be judged for example using the algorithm of Lipman-Pearson, with Ktuple:2, gap penalty:4, Gap Length Penalty:12, standard PAM scoring matrix (Lipman, D.J. and Pearson, W.R., Rapid and Sensitive Protein Similarity Searches, *Science*, 1985, vol. 227, 1435-1441).

In particular, the invention comprises a nucleic acid which encodes enzymes of SEQ ID NOS 3-17.

A preferred example of such a nucleic acid comprises SEQ ID NO 1 or a variant thereof. In particular the nucleic acid is of SEQ ID NO 1.

The term "variant thereof" in relation to a nucleic acid sequences means any substitution of, variation of, modification of, replacement of, deletion of, or the addition of one or more nucleic acid(s) from or to a polynucleotide sequence providing the resultant protein sequence encoded by the polynucleotide exhibits the similar properties as the protein encoded by the basic sequence. The term therefore includes alleleic variants, degenerate variants which encode similar proteins but vary only as a result of the degeneracy of the genetic code. It also includes a polynucleotide which substantially hybridises to the polynucleotide sequence of the present invention. Preferably, such hybridisation occurs at, or between low and high stringency conditions. In general terms, low stringency conditions can be defined as 3 x SSC at about ambient temperature to about 55°C and high stringency condition as 0.1 x SSC at about 65°C. SSC is the name of the buffer of 0.15M NaCl, 0.015M tri-sodium citrate. 3 x SSC is three times as strong as SSC and so on.

Typically, variants have 65% or more of the nucleotides in common with the polynucleotide sequence of the present invention, more typically 70%, preferably 75%, even more preferably 80% or 85% and, especially preferred are 90%, 95%, 98% or 99% or more identity.

Variants may comprise the basic sequence which has been modified to ensure that the codon usage is enhanced or optimised, as would be understood in the art, for a particular organism in which it is required that the sequence is expressed in a desired organism, for example a prokaryotic cell such as *E. coli*. This may involve modifying the content of particular nucleotides, for instance changing the percentage of G and C present in the sequence, to suit that usually found in genes which are highly expressed in the target organism. In addition, particular variants of SEQ ID NO 1 are synthetic variants, engineered to remove codons rarely found in highly expressed

genes from common expression hosts such as *E. coli* and, at the same time, avoid the introduction of codons rarely found in genes coding for O-antigens. For example, wherever possible the codons for the amino acids arg, leu, ile, gly and pro are changed to CGT or CGC (arg), CTG, CTT or CTC (leu), ATC or ATT (ile), GGT or GGC (gly), and CCG CCA or CCT (pro), thus eliminating rare codons.

When comparing nucleic acid sequences for the purposes of determining the degree of identity, programs such as BESTFIT and GAP (both from Wisconsin Genetics Computer Group (GCG) software package). BESTFIT, for example, compares two sequences and produces an optimal alignment of the most similar segments. GAP enables sequences to be aligned along their whole length and fins the optimal alignment by inserting spaces in either sequence as appropriate. Suitably, in the context of the present invention when discussing identity of nucleic acid sequences, the comparison is made by alignment of the sequences along their whole length.

SEQ ID NO 1 comprises a series of genes which encode a number of enzymes which are shown hereinafter in Figure 5 and SEQ ID NOs 3-17. Preferably any variants of SEQ ID NO 1 encode enzymes of SEQ ID NOS 3-17 or modifications of these.

The expression "fragment" used in relation to amino acid sequences refers to any portion of the given amino acid sequence which has the same activity as the complete amino acid sequence. Fragments will suitably comprise at least 20 and preferably at least 50 consecutive amino acids from the basic sequence.

The term "fragments" is also used in relation to nucleic acid sequences. Fragments of SEQ ID NO 1 may have applications in diagnostics, and these form a further aspect of the invention. For diagnostic purposes, fragments may be quite short, for example from 5-30 bases, which may be used as primers or probes. Particular characteristic regions of SEQ ID NO 1 from which suitable fragments for diagnostic purposes may be identified are elucidated hereinafter. Fragments which are

useful in therapy would generally be expected to be longer, for example from 600-17,000 bases long.

A region of genome of the *F. tularensis* strain Schu 24 (subspecies *tularensis*) which includes SEQ ID NO 1, and which is responsible for expression of the set of enzymes necessary for constructing the polysaccharide, has been identified. It is shown hereinafter in Figure 6 as SEQ ID NO 41. This sequence includes a number of genes including a series of genes that encode the enzymes illustrated in Figure 5 hereinafter as SEQ ID NOS 3-20. Putative functions were applied to these genes by comparison with known sequences as illustrated in Table 1.

Table 1

SEQ ID NO	*F. tularensis* protein	Gene product size (aa)	Putative function
2	Transposase	247	Hypothetical protein Transposase
3	WbtA	578	Sugar epimerase
4	WbtB	205	Galactosyl transferase Glycosyl transferase
5	WbtC	263	UDP-glucose 4-epimerase
6	WbtD	363	Sugar transferase
7	WbtE	436	LPS biosynthesis Dehydrogenase
8	WbtF	323	C 4-epimerase
9	Wzy	409	Membrane protein / O-antigen polymerase
10	WbtG	366	Transferase
11	WbtH	628	Asparagine synthetase
12	WbtI	360	Sugar transaminase / perosamine synthetase
13	WbtJ	241	Formyl transferase
14	Wzx	495	o-antigen flippase
15	WbtK	286	Glycosyl transferase

SEQ ID NO	*F. tularensis* protein	Gene product size (aa)	Putative function
16	WbtL	294	Glucose-1-phosphate thymidylyltransferase
17	WbtM	348	dTDP-D-glucose 4,6-dehydratase dTDP-D-glucose 4,6-dehydratase
18	Transposase	126	Transposase
19	ManC	468	Mannose-1-phosphate guanylyltransferase
20	ManB	494	phosphomannomutase

In particular the proteins illustrated as SEQ ID NOS 3-17 are believed to be involved in O-antigen biosynthesis. The O-antigen itself has applications both in diagnostics and as a prophylactic or therapeutic vaccine.

When the nucleic acids of the invention are expressed together in a host cell, they will result in the construction of an antigen that produces an immune response in an animal including a human, which is protective against infection by *F. tularensis*. Thus they may be used in the production of prophylactic or therapeutic vaccines.

The nucleic acid may be included in a vector such as a live viral vaccine, for instance, adenovirus vector or vaccinia, or in a plasmid to form so-called "naked DNA" vaccines, or preferably in a bacterial vector such as attenuated *Salmonella* species. In this case, the nucleic acid will be under the control of suitable control elements such as promoters, signal sequences, enhancers and the like, as would be understood in the art. In this case, the nucleic acid is expressed either within the cells of the patient to whom the vaccine is administered, or in the case of bacterial vectors, within the host cell itself. As a result a series of enzymes are produced which are able to construct the protective O-antigen in situ.

The vector is suitably combined with a pharmaceutically acceptable carrier in a vaccine formulation. The nature of the carrier depends upon the type of vector being used, as would be understood in the art. In particular, when the vaccine comprises a recombinant *Salmonella*, it is suitably in the form of a composition which is suitable for oral administration.

Alternatively, the nucleic acid may be included in an expression vector which is used to transform a host cell. Suitable host cells are prokaryotic or eukaryotic cells, but are preferably prokaryotic cells such as *E. coli*. In particular, the nucleic acid used is a synthetic variant of SEQ ID NO 1, optimised for expression in the particular host cells. The protective O-antigen can then be recovered from these cells after culture thereof.

Thus in a further aspect there is provided a method of preparing a prophylactic or therapeutic vaccine, which method comprises transforming a host cell with a nucleic acid of the invention, culturing said host cell, and recovering a moiety which produces a protective immune response against *F. tularensis* therefrom.

Expression vectors and host cells for use in this method, together with the product thereof form further aspects of the invention.

Vaccines of this type will suitably be in the form of a pharmaceutical composition, in which the antigen is combined with a pharmaceutically acceptable carrier, as would be understood in the art.

The compositions of the invention may be in a form suitable for oral use, for administration by inhalation (for example as a finely divided powder or a liquid aerosol), for administration by insufflation (for example as a finely divided powder) or for parenteral administration (for example as a sterile aqueous or oily solution for intravenous, subcutaneous, intramuscular or intramuscular dosing or as a suppository for rectal dosin.

The compositions of the invention may be obtained by conventional procedures using conventional pharmaceutical excipients, well known in the art.

For further information on Formulation the reader is referred to Chapter 25.2 in Volume 5 of Comprehensive Medicinal Chemistry (Corwin Hansch; Chairman of Editorial Board), Pergamon Press 1990.

The amount of active ingredient that is combined with one or more excipients to produce a single dosage form will necessarily vary depending upon the host treated and the particular route of administration. For example, a formulation intended for oral administration to humans will generally contain, for example, from 0.5 mg to 2 g of active agent compounded with an appropriate and convenient amount of excipients which may vary from about 5 to about 98 percent by weight of the total composition. Dosage unit forms will generally contain about 1 mg to about 500 mg of an active ingredient. For further information on Routes of Administration and Dosage Regimes the reader is referred to Chapter 25.3 in Volume 5 of Comprehensive Medicinal Chemistry (Corwin Hansch; Chairman of Editorial Board), Pergamon Press 1990.

The size of the dose for therapeutic or prophylactic purposes of moiety of the invention will naturally vary according to the age and sex of the animal or patient and the route of administration, according to well known principles of medicine.

Thus in a further aspect the invention provides recombinant O-antigen of *F. tularensis* which is obtainable from a host cell which expresses proteins of SEQ ID NO 3-17, or modifications thereof.

Furthermore, the applicants' realisation of the sequence of the O-antigen sequence provides the possibility that this sequence can form the basis of diagnostic tests, to determine whether a patient has an *F. tularensis* infection. In such case, samples such as blood or saliva samples may be taken from the

patient and the presence of SEQ ID NO 1 or variants thereof detected.

Specific detection methods are well known in the art, and may include amplification procedures such as the polymerase chain reaction (PCR) and/or other detection methods using for example labelled probes that hybridise to the target sequence and particularly SEQ ID NO 1. Primers and probes of this kind form a further aspect of the invention.

By selection of particular primers and probes, it may also be possible to allow differentiation between strains of *F. tularensis* infection. For instance, the applicants have found that primers comprising SEQ ID NOS 21 and 22 and 35 and 36 set out hereinafter will allow distinction between strains of *F. tularensis* subspecies *tularensis*, and *F. tularensis* subspecies *holarctica* as described below. In the former case, this possibility arises because of differences in the downstream sequence, and in the latter, because of differences in deletions in the flanking transposase sequence. Consequently, analysis using primers or probes based upon these regions may be used to determine whether any particular strain is *F. tularensis* subspecies *tularensis* or otherwise.

In order to discover whether the LPS from a subspecies *tularensis* strain has similar structure (and properties) to that from a subspecies *holarctica* strain, LPS from *F. tularensis* strain Schu S4 (subspecies *tularensis*) was extracted.

LPS extracted from *F. tularensis* strain Schu S4 was shown to have a characteristic ladder pattern after gel electrophoresis. However, the LPS was difficult to stain and required additional oxidation in order to visualise the O-antigen bands. This may suggest that the sugars in the O-antigen of *F. tularensis* strain Schu S4 are not oxidised in the same way as the O-antigen sugars found in most other bacteria.

The *F. tularensis* strain Schu S4 O-antigen gene cluster contained 15 genes, the putative functions of which was assigned (see Table 1 above) based on the BLAST results and structural information about the sugars contained in the O-antigen. Genes

within the cluster are likely to be responsible for the production of the O-antigen molecule as well as the transportation of the molecule out of the bacterial cell.

There are two main O-antigen synthesis modes, O-antigen polymerase (wzy)-dependent and wzy-independent. In the wzy-dependent system it is thought that the polymerase (wzy), flippase (wzx) and chain length determinant (wzz) are part of a complex in the cell wall which facilitates polymerisation and export of the LPS molecule. In the wzy-independent system a different set of proteins are involved in the transportation and polymerisation of the LPS molecule. The transporter is ATP driven and composed of two proteins wzt and wzm that belong to the ABC-transporter family.

In the *F. tularensis* O-antigen gene cluster, proteins with high identity to wzy and wzx are present, suggesting that transportation and polymerisation of the O-antigen is via a wzy-dependent pathway.

The TMHMM analyses of the putative O-antigen flippase (Wzx) and polymerase (Wzy) proteins supported their assigned functions based on sequence similarity. The predicted numbers of trans-membrane helices for the *F. tularensis* proteins of 14 and 11 for Wzx and Wzy respectively are similar to those predicted for other bacteria, in which these cytoplasmic membrane proteins have been predicted to have around 10-12 trans-membrane helices. The prediction of 2 large periplasmic domains for the *F. tularensis* Wzy protein is consistent with the two large periplasmic domains of the *Shigella flexneri* Wzy protein.

No gene that could encode a Wzz homologue was identified, which may indicate that one is not present in the *F. tularensis* genome.

The close proximity or overlapping of the genes *wbtA* to *wbtL* suggests these are transcribed as one operon.

Approximately 0.5Kb downstream is wbtM, which has a putative promoter of its own. Downstream of the second transposase are *manC* and *manB*, which also have their own putative promoter and are probably not involved in biosynthesis

of the O-antigen as mannose was not found to be part of the structure of the *F. tularensis* O-antigen, nor is it one of the intermediate products required for its synthesis.

The two genes *manC* and *manB* may once have been involved in biosynthesis of the O-antigen in an ancestor of *F. tularensis*. The presence of transposases flanking the O-antigen biosynthetic gene cluster *wbtA* to *wbtM* suggests this cluster may have been horizontally acquired, perhaps replacing an ancestral polysaccharide gene cluster.

The O-antigen gene cluster appears to be present in all *subspecies tularensis* and B strains screened. However, there is at least one difference between the clusters in *subspecies tularensis* and B strains within a region containing a transposase. BLAST analysis using the partially deleted transposase has revealed possibly over 50 copies of it in the *F. tularensis* Schu S4 genome. It is possible that the insertion sequence originated in the *F. tularensis* genome from *S. pneumonia* and was copied randomly within the genome. The open reading frames flanking the insertion sequence have no significant homology within the *F. tularensis* genome, suggesting that these genes were not imported to this locus with the insertion.

In subspecies *tularensis* strains, this insertion has become deleted to leave only fragments of the transposase and downstream sequence. The overall similarity between the subspecies *tularensis* and subspecies *holarctica* clusters seems to indicate that the insertions took place in *F. tularensis* before division of the subspecies. Partial deletion of the *subspecies tularensis* transposase would have the effect of stabilising this region of DNA, as this enzyme has been found to be necessary for insertion events to take place.

It seems unlikely that this will affect expression of the cluster in either subspecies *tularensis* or B strains. It could be that in subspecies *tularensis* strains part of the transposase has been lost due to genome down sizing. However, the gross difference in size of PCR products generated across this region

when amplifying DNA from different subspecies may be utilised in diagnostic procedures.

The applicants have found that a similar O-antigen gene cluster to that found in *F. tularensis* strain Schu S4 is present in other strains of *F. tularensis*. This includes subspecies *holarctica* strains. Consequently, a vaccine which utilises the O-antigen to produce a protective immune response is likely to provide protection against infection by several virulent strains of *F.tularensis*.

The applicants have demonstrated that LPS from *F. tularensis* subspecies *tularensis* strains is protective. In particular, it appears to be protective against challenge from strains other than *F. tularensis* subspecies *tularensis*, and in particular against challenge with *F. tularensis* subspecies *holartica*. This finding is unexpected in view of the results reported above which suggest that LPS from *F. tularensis* subspecies *holartica* is not protective against infection by other *F. tularensis* species. Thus recombinant vaccines as described above will be particularly useful.

Thus in a further aspect, the invention provides LPS obtainable from *F. tularensis* subspecies *tularensis* for use as a vaccine against infection by *F. tularensis*. Vaccine compositions containing LPS from *F. tularensis* subspecies *tularensis* are also novel and form a further aspect of the invention. These will comprise pharmaceutically acceptable carriers as described above.

The invention will now be particularly described by way of example with reference to the accompanying diagrammatic drawings in which:

Figure 1. SDS-PAGE analysis of LPS isolated from *E. coli* strain K325, 1.25 μg (track 1) and *F. tularensis* strain Schu S4, 50 μg (track 2).

Figure 2. The genetic organisation of the O-antigen gene cluster in *F. tularensis* strain Schu S4. The G+C content of the O-antigen cluster is shown in the upper panel.

Figure 3. Schematic structure of an O-antigen subunit of *F. tularensis* strain Schu S4 and the assignment of putative functions to the O-antigen gene cluster genes. A single O-unit is shown with sugar residues and glycosidic linkages indicated.

Figure 4. Shows the region of the genome the nucleic acid sequence of the F. tularensis genome which encodes all the proteins shown in Figure 5.

Figure 5. Shows the amino acid sequences of proteins encoded by SEQ ID NO 1, as well as a number of flanking gene sequences,

Figure 6. Shows the nucleic acid sequence (SEQ ID NO 1) which encodes the enzymes necessary for O-antigen production.

Example 1

Methods

Bacterial strains and growth conditions

Bacterial strains used in this study are shown in Table 2 and were cultured at 37°C on BCGA agar for 48 hrs.

Table 2

Species and Strain	Subspecies
F. tularensis Schu4	tularensis
F. tularensis 199	tularensis
F. tularensis 230	tularensis
F. tularensis 041	tularensis
F. tularensis LVS	holarctica
F. tularensis 200	holarctica
F. tularensis 025	holarctica
F. tularensis 075	holarctica
F. tularensis HN63	holarctica
F. tularensis 147	mediaasiatica

LPS purification

LPS was purified from *F. tularensis* strain Schu S4 using a hot-phenol and water extraction method (Westphal O, et al. (1965). *Methods in Carbohydrate Chemistry* **5**: 83-91).

Gel electrophoresis and silver staining

Glycine gel electrophoresis was performed according to the method of Laemmli (Laemmli UK. (1970). *Nature* **227**: 680-685.) using a 12.5 % separating gel with a 4.5% stacking gel. Ten µl of each sample were electrophoresed for approx 2 h at 100 mV in the Mini-protean II slab system (Biorad).
Gels were silver stained according to the method of Chart (Chart H. (1994) LPS: Isolation and Characterisation. In: Raton B, Arbor A (eds.) *Methods in Practical Laboratory Bacteriology*. CRC Press, London,Tokyo, pp. 11-20). However, the oxidation step was increased to 10 min.

Nucleotide sequence analysis

The sequence encoding the O-antigen biosynthetic cluster was identified and extracted from the Known protein sequences (obtained from GenBank) involved in the biosynthesis of the O-antigen of other bacteria were used to probe the *F.tularensis* Schu S4 partial genome sequence (Prior RG, et al. (2001) *Journal of applied microbiology* **91**: 614-620), available at http://artedi.ebc.uu.se/Projects/Francisella/, using TBLASTN (Altschul SF. et al. (1997) *Nucleic acids research* **25**: 3389-3402). The contig containing the putative O-antigen gene cluster was extracted and subsequently analysed using the annotation tool Artemis (http://www.sanger.ac.uk/Software/Artemis). This allowed visualization of BLASTN, BLASTX and BLASTP searches, GC content and other analyses performed on the sequence and the predicted proteins.

The protein sequences encoded by the putative O-antigen flippase gene (*wzx*) and O-antigen polymerase gene (*wzy*) were analysed for trans-membrane helices using TMHMM (Sonnhammer ELL, et al. (1998) In: Glasgow J, Littlejohn T, et al. (eds.) *The*

sixth international conference on intelligent systems for molecular biology. AAAI Press, Menlo Park, CA, pp. 175-182).

PCR analysis of the putative O-antigen gene cluster

DNA was prepared from the *F. tularensis* strains shown in table 1, by phenol extraction, as described by Karlsson *et al* 2000 (*Microb.Comp.Genom.* **5**: 25-39). Ten pairs of overlapping PCR primers were designed to amplify the whole of the putative O-antigen gene cluster in approximate 2 kilobase segments using the DNAstar program PrimerSelect. The primers were designed with annealing temperatures ranging from 42 to 59°C, although all were used successfully at 49°C.

The structures of these primer pairs is summarised in Table 3.

Table 3

Primer set	Forward/ reverse	Structure	SEQ ID NO
1	Forward	ATAATGAAATCAATCCACGAG	21
	Reverse	CCAGCCAGTCAGTCCCACAG	22
2	Forward	TGTCTTAGATATGGGGCAACC	23
	Reverse	ACAAATATCAAATCCTAACACATC	24
3	Forward	TAGAAGCAGCTGCGATAGGTAGAC	25
	Reverse	TTAAATAAAAACTGAGGAAACA	26
4	Forward	ATGGTATTTTAATCAAGTGT	27
	Reverse	CTAGTATGCCCCAGAGT	28
5	Forward	TGGTGCGACAATCAAGTTA	29
	Reverse	AGAAGTTCCTCCTCAGTC	30
6	Forward	AGAAATTAAGAGCAAAAGGAAAGT	31
	Reverse	ATCTCAAAGTCAAAATCAGTCTCT	32
7	Forward	TACGATATTGTCCTCTCCGATTAG	33
	Reverse	TAGTTGCGACATATTGACCTG	34
8	Forward	AGGCAGGTCAATATGTCGCAACT	35
	Reverse	TTTCCGCAACACTTCAGCAACTT	36
9	Forward	GCTATGGCCACTATCACGAGAGG	37
	Reverse	TATACTTGCTTGCCCACTGCTTAG	38
10	Forward	ACCGTAGTGAGCATTGGATTGT	39

Primer set	Forward/ reverse	Structure	SEQ ID NO
	Reverse	ACTAGGGCCTCTGACCGTTCTC	40

PCR amplification using each pair of primers with each template DNA was carried out in the following mixture: 1x PCR buffer (including 1.5 mM $MgCl_2$), 0.2 mM deoxynucleoside triphosphates (dNTPs), 2.5 mM forward primer, 2.5 mM reverse primer, 2.0 µl template DNA, 0.5U *Taq* polymerase and filtered sterile water to a final volume of 20 µl. The reaction mixtures were incubated at 90°C for 1 min and then cycled at 90°C for 1 min, 49°C for 1 min and 72°C for 2 min 25 s for 30 cycles, with a final incubation at 72°C for 10 min. PCR products were visualised on 0.5% agarose gels, with ethidium bromide staining. PCR buffer, dNTPs, and polymerase were from Roche. PCR primers were synthesised by MWG-Biotech.

Cloning of PCR Products

PCR products amplified from Schu S4, HN63 and LVS DNA using primer pair 8 were cloned into pGEM-T easy (Promega) for sequence analysis. Ligated DNA was transformed in *E. coli* JM109 chemically competent cells (Promega) and putative clones were screened using both colony PCR and digestion with restriction endonucleases. All DNA manipulations, including ligations, transformations, colony PCR, restriction endonuclease digestions and agarose gel electrophoresis were carried out according to methods described by Sambrook *et al* (1987) Molecular cloning: A laboratory manual. Cold Spring Harbor, New York).

Purification of PCR products from agarose gel was achieved using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions.

The three constructs were sequenced at Oswel by the dideoxynucleotide chain-termination method (Sanger F, et al. *Proc.Natl.Acad.Sci.U.S.A.* **74**: 5463-5467) using universal primers. Each sequence was compared and the BLAST (Altschul SF, et al. (1990) Basic local alignment search tool. *J.Mol.Biol.* **215**: 403-410) function of the ARTEMIS software package was used

for homology searches in the locally held GenBank databases to identify the functions of the differential regions of DNA.

Mass spectrometry analysis of the O-antigen molecule

Results

LPS purification

The hot phenol-water extraction method was used to purify LPS from 2.2 g of freeze dried *F. tularensis* strain Schu S4. This resulted in 7 mg of LPS, which is a yield of 0.3 %. The LPS was difficult to visualise after SDS-PAGE and silver staining. The oxidation step was increased from 5 min to 10 min to visualise a ladder pattern (Fig 1).

***F. tularensis* O-antigen biosynthetic gene cluster**

The *F. tularensis* O-antigen biosynthetic gene cluster was found to be 17Kb in length and contain 15 genes putatively identified as being involved in O-antigen biosynthesis, flanked by two transposases (Fig 2). Possible promoter sites were identified just upstream of the genes *wbtA* and *wbtM*. Downstream of the second transposase are located the genes *manC* and *manB*, with a possible promoter just upstream of *manC*.

Figure 2 also shows the G + C content plot of the cluster using a window size of 500 bases. The overall G + C content of this region of the genome at 31.27% is slightly lower than the genome average of approximately 33%. The plot shows that the central section of the cluster, from *wzy* to *wbtK*, generally has an even lower G + C content.

Downstream from *manC*, on the opposite strand, are located the genes for the transcription termination factor rho and thioredoxin. In *E. coli* both of these genes are also found flanking one end of a polysaccharide biosynthetic gene cluster - that of the enterobacterial common antigen.

The O-antigen repeat unit of *F. tularensis* is shown in Fig 3, together with the putative role of the genes involved in O-antigen biosynthesis. Based on their homology to other LPS and sugar biosynthetic genes, in particular *P. aeruginosa* serotype O6 which expresses a similar O-antigen repeat structure (Knirel

YA, et al. (1985) *Eur.J.Biochem.* **150**: 541-550), the putative role of the gene products have been assigned.

It is proposed that the biosynthesis of 2-acetamido-2,6-dideoxy-D-glucose (QuiNAc) involves WbtA, a dehydratase and WbtC, which shows homology to UDP-Glc 4-epimerases. WbtA and WbtC share homology to WbpM and WbpV of *P. aeruginosa* strain O6, both thought to be involved in QuiNAc biosynthesis and shown to be essential for O6 O-antigen synthesis. WbtE, WbtF and WbtH are proposed to be involved in 2-acetamido-2-deoxy-D-galactouronamide (GalNAcAN) biosynthesis. WbtF shows homology to UDP-glucose 4-epimerases, including WbpP and VipB, whilst WbtE shows homology to WbpO and VipA, UDP-GalNAc dehydrogenases involved in the formation of 2-acetamido-2-deoxy-D-galactouronic acid (GalNAcA) in *P. aeruginosa* and *Salmonella enterica* var *typhi* respectively. WbtH produces significant alignments with glutamine amidotransferases, including WbpS of *P. aeruginosa* serotype O6, which may putatively be involved in the formation of the GalNAcAN amido group. Biosynthesis of the fourth sugar, 4N-formyl-quinovosamine (Qui4NFm) most likely involves WbtI, WbtJ, WbtL and WbtM. Sequence homology suggests that WbtL may be involved in the formation of the activated sugar dTDP-D-Glucose with WbtM functioning as a dTDP-D-Glucose 4,6-dehydratase. WbtI is proposed to be involved in Qui4NFm amination since it shows homology to RfbE, a perosamine synthetase. Finally, WbtJ is likely to be responsible for the addition of the N-formyl moiety, showing significant homology to formyltransferases.

Specific glycosyltransferases are required to form the oligosaccharide units of the O-antigen repeat. Four glycosyltransferases would be necessary for the synthesis of each O-antigen unit in *F. tularensis*. Based on homology, WbtB is proposed to mediate the addition of QuiNAc to undecaprenyl phosphate (Und-P) to initiate O-antigen biosynthesis. WbtD and WbtG are probable GalNAcAN transferases, possibly involved in the addition of the two consecutive GalNAcAN residues onto the O-antigen unit. WbtD shares homology to WbpU of *P. aeruginosa*

strain O6, proposed to transfer 2-formamido-2-deoxy-D-galactouronamide (GalNFmAN) onto QuiNAc (Belanger M, et al. (1999). *Microbiology* **145**: 3505-3521). WbtG is homologous to WbpT of *P. aeruginosa*, thought to be involved in addition of GalNAcA to GalNFmAN. WbtK is probably the fourth glycosyltransferase, which adds 4,6-dideoxy-4-formamido-D-glucose (QuiNA4Fm) to complete the tetrasaccharide O unit.

Wzx and Wzy

Once assembled, the O-antigen repeat units are translocated to the periplasmic face of the inner membrane via Wzx, a transporter/flippase. Wzy then acts as an O-antigen repeat unit polymerase. When analysed using TMHMM, the *F. tularensis* Wzx protein had a predicted 14 trans-membrane helices, with both termini on the cytoplasmic side of the membrane. The *F. tularensis* Wzy protein had a predicted 11 trans-membrane helices, with the amino terminus predicted to be on the cytoplasmic side of the membrane, and the carboxy terminus on the periplasmic side. Additionally, the Wzy protein was predicted to have two large periplasmic domains from amino acids 142-178 and 268-327.

A gene with homology to the O-antigen chain length determinant (*wzz*) was not identified in the current *F. tularensis* Schu S4 sequence dataset.

PCR analysis of the O-antigen gene cluster

Eight of the PCR products (primer sets 2,3,4,5,6,7,9 and 10) from each template DNA appeared to bethe same size when viewed by agarose gel electrophoresis. Primer pair 1, covering the start of the gene cluster, had to be designed to amplify a 4.8 Kb region due to lack of suitable priming sites upstream of the cluster because of the presence of an insertion element found many times in the *F. tularensis* Schu S4 genome. This primer pair 1 produced the relevant size product for *F. tularensis* Schu S4, but when used on subspecies *holarctica*, strain LVS, did not produce a product. Thus this primer pair may have particular applications in diagnostics where distinction between *F. tularensis* subspecies *tularensis* and F.

tularensis subspecies holarctica is required. Where samples containing DNA from the former is present, a PCR using primer pair 1 will generate a product, which would not be present in the second case.

The PCR using primer pair 8 revealed a difference in size between subspecies *tularensis* strains and subspecies *holarctica* and subspecies *mediaasiatica*. Subspecies *tularensis* strains show a deletion of 303 nucleotides when compared to subspecies *holarctica* strains (including LVS) and subspecies *mediaasiatica*. Cloning and sequence analysis of this region from the subspecies *tularensis* strain Schu S4, the subspecies *holarctica* strain HN63 and LVS has shown that the deletion in Schu S4 occurs at the beginning of a putative transposase that is similar to IS630-spn 1 transposase ORF 1 of *Streptococcus pneumoniae*.

Thus primer pair 8 may also be particularly useful in distinguishing between strains of *F. tularenis*. Following a PCR reaction on samples containing DNA using these primers, a separation of the products on the basis of size, for example on a gel, should reveal distinguishable differences therebetween.

Example 2

Protective Effects

LPS purification

LPS was purified from *F. tularensis* strain Schu S4 or from strain LVS using a hot-phenol and water extraction method mentioned above in Example 1.

Immunization with LPS and protection studies

The ability of *F. tularensis* strain LVS or strain Schu S4 LPS to protect BALB/c mice from a *F.tularensis* was determined by immunizing groups of six female BALB/c mice by the i.p. route with the purified LPS obtained. On each dosing occasion, mice were given 50 μg of LPS in phosphate buffered saline (PBS). The mice received three immunizations, each 7 days apart.

Mice were challenged i.p. with *F. tularensis* LVS (1 x 10^5 CFU) 21 days after the last immunization. All control animals died after

challenge. Mice which had been immunised LPS isolated from the LVS strain were protected from death. Mice which had been immunised with LPS from either the SchuS4 or LVS strain showed and extended time to death. At a challenge dose of 10 cfu animals immunised with SchS4 LPS survived for an average of 64 hours (with 99 % confidence) longer than the unimmunised controls.

Claims

1. A nucleic acid which encodes a series of enzymes or enzyme fragments which, when expressed together in a cell, are able to produce an immunogenic moiety able to produce an immune response in an animal to which it is administered, which response is protective against *Francisella tularensis* infection, said nucleic acid encoding at least some enzymes of SEQ ID NOS 3-17 or modifications of these.

2. A nucleic acid according to claim 1 which encodes enzymes of SEQ ID NOS 3-17.

3. A nucleic acid according to claim 1 or claim 2 which comprises SEQ ID NO 1 or a variant thereof.

4. An nucleic acid according to claim 3 which is of SEQ ID NO 1.

5. A nucleic acid according to any one of the preceding claims wherein the codons have been optimised for expression in a bacterial cell.

6. A nucleic acid according to claim 5 wherein the bacterial cell is *E. coli*.

7. A nucleic acid comprising a fragment of SEQ ID NO 1 which may be used to detect the presence of SEQ ID NO 1 in a sample.

8. A nucleic acid according to claim 7 which comprises an amplification primer.

9. A nucleic acid according to claim 8 which is selected from SEQ ID NO 21, 22, 35 or 36.

10. A live vaccine vector, which comprises a nucleic acid according to any one of claims 1 to 5.

11. A live vaccine vector according to claim 10 which comprises a bacterial vector.

12. A live vaccine vector according to claim 11 wherein the bacteria is a *Salmonella* species.

13. A vaccine comprising a live vaccine vector according to any one of claims 10 to 12 in combination with a pharmaceutically acceptable carrier.

14. A method of preparing a prophylactic or therapeutic vaccine, which method comprises transforming a host cell with a nucleic acid according to any one of claims 1 to 6, culturing said host cell and recovering a protective immunogenic moiety from the culture.

15. An expression vector comprising a nucleic acid according to any one of claims 1 to 6.

16. A host cell transformed with a vector according to claim 15.

17. Recombinant O-antigen of *F. tularensis* obtainable by a process according to claim 14.

18. A vaccine comprising recombinant O-antigen according to claim 17 in combination with a pharmaceutically acceptable carrier.

19. A method of diagnosing infection by *F. tularensis* infection, which method comprises detecting in a sample taken from a patient suspected of having an infection a nucleic acid sequence according to claim 7.

20. A method of differentiating between strains of F. tularensis, which method comprises selecting primers or probes which are specific for SEQ ID NO 1, and not for similar sequences in subspecies other than *F. tularensis* subspecies *tularensis*, or which produce distinguishable products when used to analyse other species, and conducting an analysis using the said primers or probes.

21. A method according to claim 20 wherein the analysis is conducted using a polymerase chain reaction (PCR) and a pair of primers.

22. A method according to claim 21 wherein the primers are specific for a start region of SEQ ID NO 1.

23. A method according to claim 22 wherein the primers are of SEQ ID NO 21 and SEQ ID NO 22.

24. A method according to claim 23 wherein the primers are specific for the end transposase coding region of SEQ ID NO 1.

25. A method according to claim 24 wherein the primers are of SEQ ID NO 35 and SEQ ID NO 36.

26. Lipopolysaccharide (LPS) obtainable from *F. tularensis* subspecies *tularensis* for use as a vaccine against infection by *F. tularensis*.

27. LPS according to claim 26 where the strain of F. tularensis subspecies tularensis is the Schu4 strain.

28. A pharmaceutical composition comprising LPS according to claim 26 or claim 27 in combination with a pharmaceutically acceptable carrier.

Fig.1.

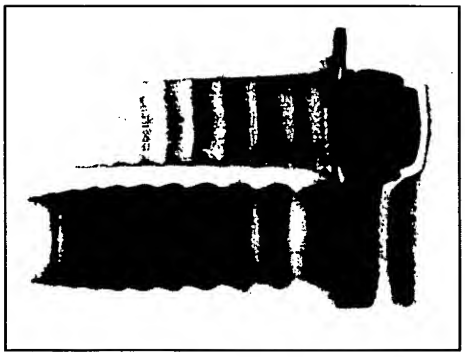
1
2

Fig.2.

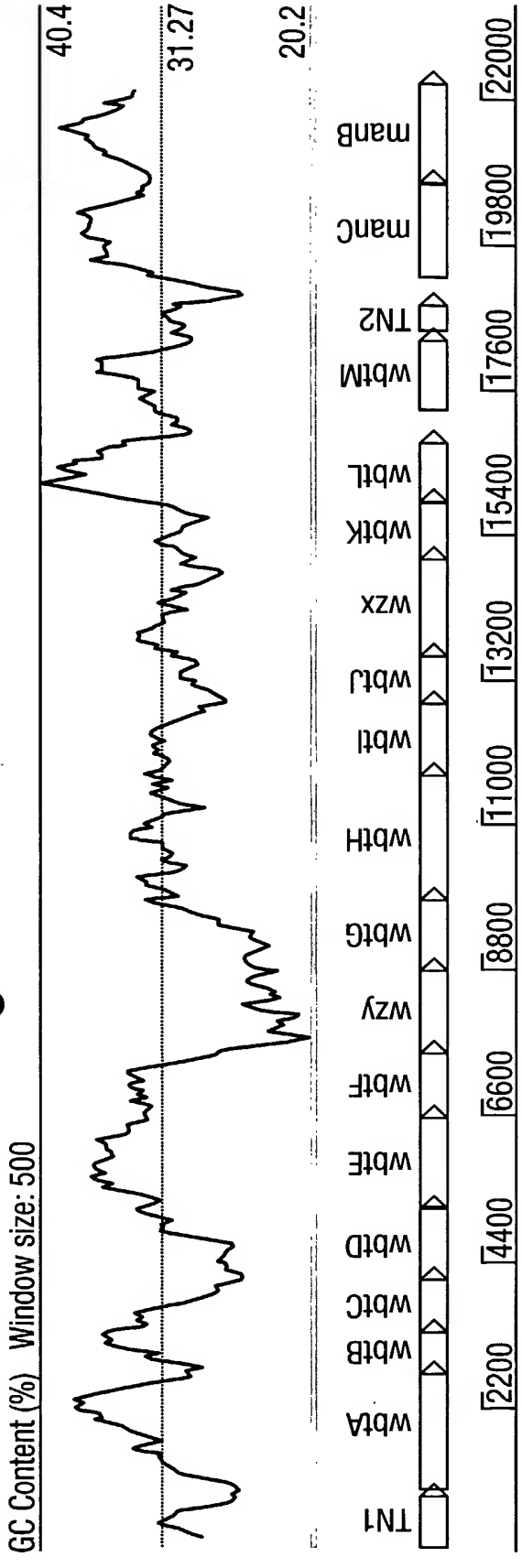
GC Content (%) Window size: 500
40.4
31.27
20.2
TN1
wbtA
wbtB
wbtC
wbtD
wbtE
wbtF
wzy
wbtG
wbtH
wbtI
wbtJ
wzx
wbtK
wbtL
wbtM
TN2
manC
manB
2200
4400
6600
8800
11000
13200
15400
17600
19800
22000

Fig.3.

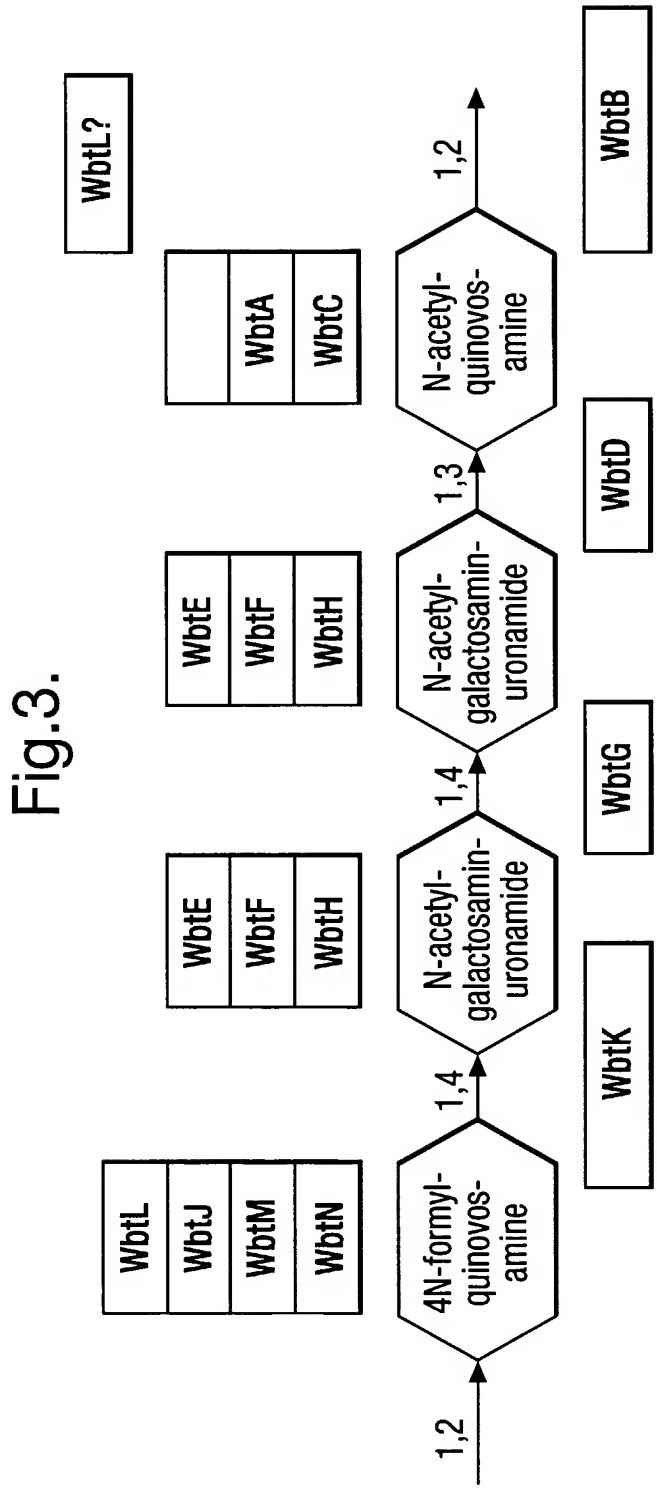
WbtL?
WbtA
WbtC
WbtB
1,2
N-acetyl-quinovos-amine
1,3
WbtE
WbtF
WbtH
N-acetyl-galactosamin-uronamide
WbtD
1,4
WbtE
WbtF
WbtH
N-acetyl-galactosamin-uronamide
WbtG
1,4
WbtL
WbtJ
WbtM
WbtN
4N-formyl-quinovos-amine
WbtK
1,2

Fig.4.

SEQ ID NO 41

```
1     tcttttataa atgatgatag caaacaaaaa ataataggtt ctgtgcacaa aaaacttaaa
61    ttatattgaa aatagctaaa ccgctgttat ttaagatttg aaaagcaata aatatcaatg
121   gtttagcaaa tgaattatca tataaaagaa gtattctggt caattatttt atcattctta
181   aaatcacaaa aaggtataca taccaatgat gaagccaaat taagattgtt tattgaagct
241   gtattttatg tgttacgtac aggctgtcaa tggagaatgt taccatttta ttatggtaaa
301   tatagatcaa tacataagcg ttttaaagat tggtgtgata aagatatatt ttctagatta
361   tttaaatcag tacaaaaccc tgatttacaa gaagtcatgc ttgattcaac aatagcaaga
421   gcacatgctt gtgctacggg atatgataaa gatgataacc aagcaattgg tagatcagtt
481   ggtaggataa ccactaaaat ccatgctatg actgatgctt taggtaatcc aatagaaata
541   ttgttgtcag aggataaaac tcatgatagt aaagtagcta tagatttact aaaaaatgta
601   tataatacaa aagttatcgc tgatagagca tatcattcta atgaaatcag gcagcatatt
661   caaggtatat cctctgaagc tgttatccct tgtaaatcaa atactctaaa ccatatacct
721   tttgatagtc atgtatataa agaaagacat ttgatagaga atttcttttc taaaattaag
781   cattttagaa gagtattctc tagatttgat aaaaccattt tagcatatat aggaatgatt
841   aaattagctt gtacttttat ttggttacga tgaatattta tttttgtgca cagaacctaa
901   tttgcatttt tgtgcacaaa gaaaattttt ttgatataat agactttaat aggatatttt
961   ctaaaaatta acaaatgtct ttctacgata atagaacgct taatttcgtg gtaataatag
1021  ttttaactat tattactgtt aattggactt tctatatttt caagcaagat gttaatttac
1081  atttttact tgcattagtt ttgctgagat gcttgtcatc ttttttacta cttagagatt
1141  atatggctag ttggcgtaag tcgactcaaa aaacttttt acgtaaggct tttattaatt
1201  tgccagtatt tttcatagtg gcattatttt tttatggcaa agtcactttt tcgttgatat
1261  tctctgagtt tttattttat gttttttga tcagtttaag tgtctacttt tattggtatt
1321  tgatgaacag aggatcagtg gataaaagta aaactgcggt tatttatggt gcaggtgctg
1381  caggaacaaa gattgctcaa gaacttgctt ctgctggtta tcgcatcaaa tgttttgttg
1441  atgacaatga aactttacaa aaaagaagta ttgatagtaa aaaggttcta tctaaagctg
1501  aattaacaaa actattgcta tctagtagat ttgacctttt ggttattgca ttgccaagaa
1561  atgcaaacca agtagtcaaa aatatatata aagaatttga aaaggatttt aatcagatta
1621  gaattatgcc gcctcttgag gaaattcttc aagatgagaa ttttatgtca cagttgaagc
1681  ctgtttcact ctatgatcta ttagcgcgtg atactaagag tttagataaa gaatctatct
1741  ctaattttat caaaaataag gtggtgctag tcacaggagc tggaggtagt ataggttctg
1801  aaatagtaca tcaatgtatc aagtatcagg caaaagagtt gatattggtt gatcatagtg
1861  agtttaactt atataaaatt actgaggagt gtagtcattt taatatcaat agtgtgctat
1921  gttctgtttg tgatagaaaa gcattggctg aggtttttca aaagtatact ccaaatatag
1981  tatttcatgc tgctgcctac aagcatgttc ccttagttga ggagaatatc tctagagcaa
2041  ttagaaataa tatcttaggt actaagaatg ctatagatct ggctatagaa gctggtgttg
2101  agtcatttat attgatttcc actgataaag cagtgcgacc aacgaatgtt atgggggcta
2161  ccaagagagt ttgtgagctg tatttacaga atgttgatcc caaaaatacc aagcttgctg
2221  cagtgcgttt tggtaatgtg cttggtagta gtggcagtgt gattccaaaa tttgaagagc
2281  aaataagaaa aggtggtcct gttacagtta ctcatcctga aattacacgt tattttatgt
2341  tgataccaga agcttgtgaa ctggtcctac aagctggtgc tattgcaaaa aattcagagg
2401  tctttgtctt agatatgggg caacctgtca agattattga tcttgctaaa caatttatta
2461  gactttctgg tagaggtgat attgatatta aaatagttgg tttgcgtcca ggagagaaac
2521  tttacgaaga gcttttgata gaggaagatg atgttagtac cgactataaa gatattttta
2581  ttggtagaag gacttttttac gatattaata ctctaaacca agatattgaa tcgttgatca
2641  aggatgatgt tgatcagctt gtgatattaa agaaaattgt tccggaattt gaacatagat
2701  tgaatgggta gtggttttat gttttatgag gtttttaaaa gattgcttga tattttactt
2761  tcttttatgg ggttgttgtt attaagtcct attttcttaa ttattatttt tatgataaag
2821  aaagattcaa aaggacctat atttttttaaa caaaagcgct atggtaaaga taagcaattt
2881  ttttacatat ataagtttag aactatgtat gttgatactc caaaagatat gccaacgcac
2941  atgttacagg atccatcgaa atgtataact aaggttggag gatttttaag gaaatcatct
3001  ttagatgagt tgccacaaat tataaatatt ctaaaaggtg aaatgagcat cgtgggtcca
3061  agaccagcat tatggaatca agatgactta atagcacaaa gagataagta tggggcaaat
```

Fig.4 (Cont I).

```
3121  gctgtgcctg tgggactgac tggctgggca cagattaatg gtagggatga attaccaata
3181  cctgataaag ctaaacttga tggtgattat gtaaaaaata aaagtacatg gtttgattta
3241  aaatgtattt ttttgacagt attttctgtt tttgccaaaa agggcgtcgt tgagggtggt
3301  actggagctt taggtaacaa agaggattta aagtagtatg aaaaaaagaa tcttagttac
3361  aggtttgagt agctatattg gtaactcatt tgcggctaaa tataactcag attttagtat
3421  cgataaaata tctttgcgcg atgtttcgtg ggcaaatata gacttaagtg gttatgatgc
3481  tgtattgcat gtcgctggaa ttgcccatac ttcaaaggat cctaaactaa aagaaaaata
3541  ctataaaata aatacgcaat taacttatga tctggcaaaa caagctaaag atcaaggtgt
3601  tcgacagttt gtgtttttaa gtagtattat agtttatggt gatagtgcgc caataggtca
3661  acaaaaagtt ataactaaat ataccgaacc taaaccagat gatttttatg gagatagtaa
3721  gcttcaaact gaaattaagc taaatagcct ggctagtgat gactttaata ttgctataat
3781  cagaccacca atggtatatg gagaaggctc aaaaggcaac tatccaaagt tggttaaact
3841  tgcaaagtat acttttattt ttcctaatat taataaccaa agaagtgtta tatctataga
3901  taatttatct aaagagattg cagaaataat tttgcaaact aaacatggag tttttctact
3961  tcaagataat gaatattttt gcacttcaca gtttataaaa aactatagaa aagatgtttt
4021  aggtaagaga acttatctga caaaaatttt taatccaatt ataagattgc ttgctaaaaa
4081  agtagatttt attaataaag tttttgggaa tttgacttat gagaagtaag ttattattca
4141  tagctaatga ttttgatatt gtaatatatc gtttcagaag agaagtaatc gagtcttttg
4201  ctgctaaaga gtatgagata gtactagtaa caccatattc taagaaagca gaggtttttt
4261  gtaaaagtct tggtgttaag tatataaatg ttgatataga tagacgaggc aaaaatcctt
4321  ttaaggattt gcttctttta tttaactatt tcaaaataat aaaaaaagaa aaacctgatt
4381  acatttttag ctatacaatt aaaccaaatt tgtatgttgg gttagtgaat ttgtttttta
4441  ggaagaagtt ttatccaaat gtaacaggct taggaagtgt ttttgctaat catggtattg
4501  ttcagaagtt tataatatct ttatataagt tatcatttaa aagcaccaca aaagtattct
4561  ttcagaatga gcaaaataaa aagttattta tagctaagaa aataatcagt ggagaaaaat
4621  caatattatt accaggttct ggggtaaact tagatgaaaa taaatatgtt gactatccta
4681  aagaccaagg aatattaaaa ttcgtttttc ttggccgaat aatgaaagaa aaggggattt
4741  atgaattgtt agaagccttt gctatacttg agaaaaaata taaaaatatt agtcttgaca
4801  tttatggttt ttgtgatgaa aataaatcta attttatggg aaaggttaat acgataaaat
4861  cagtaaaatt ttatggtttt actgataata ctaaagaaaa aatagctagt gcacatgcag
4921  ttgttttgcc atcttaccat gaaggaatgt caaatgtgct gttagaagca gctgcgatag
4981  gtagacctgt aattgcgtca gatattcctg ggtgtagaga aatttttgat gatggtctct
5041  ctggcttatc atgtaaccct aatgatgtga gttctttacg taactcatta gagcagttta
5101  taaatatgtc gtatactgat aaaatagcta tgagctataa agctagagct aagatagaaa
5161  aagattttga tagaagtatt gttgtcaatg catacttaca gcaaaattaa taataagggt
5221  ttaaattatg agtttatatg aggatatagt cgctaaaaga gaaaaggttt cattggttgg
5281  cttgggttat gttggtttac caatagctat tgcatttgca aaaaaaatag atgtgttagg
5341  atttgatatt tgtgaaacaa aagttcaaca ttataaggat ggttttgatc caacaaaaga
5401  agtaggagat gaggctgtca gaaatacgac aatgaaattt agttgtgatg aaacaagtct
5461  taaagagtgt aaatttcata ttgttgcagt tcctacacca gttaaagcag ataaaactcc
5521  tgatttgacg ccgattatta aggcaagtga gacggttggt aggaatcttg tcaaaggcgc
5581  ttatgttgtg tttgaatcaa ctgtttatcc tggtgttaca gaagatgttt gcgtaccaat
5641  acttgaaaaa gagtctggct tgaggtctgg tgaagatttc aaagttggtt actctcctga
5701  gaggataaat cctggtgata aggttcatag gttagaaaca attatcaaag tagtatctgg
5761  tatggatgaa gagtctttag atactatagc aaaagtttat gagctagtag tagacgcagg
5821  agtttataga gctagtagta taaaagtggc tgaagctgct aaggttatag aaaactctca
5881  aagagatgtt aatatagctt ttgttaatga gttatcgata atatttaatc agatgggtat
5941  tgatactcta gaggttttag cagcagctgc aactaaatgg aatttcttaa actttaagcc
6001  tggtcttgtt ggtggacatt gtattggtgt tgacccatat tacctaacgt acaaggcagc
6061  tgagcttgga tatcattctc aggtaatatt atctggtcgt aggataaatg atagtatggg
6121  taaatttgta gttgagaatt tagtcaaaaa actgatatct gcagatatac ctgttaagcg
6181  agctagagta gcaattttcg gctttacttt taaagaagac tgtcctgaca ctaggaatac
6241  tcgagttata gatatggtaa aagagctcaa cgagtatggt atagagccat atattataga
6301  tccggtagct gataaagaag aggctaaaca tgagtatgga cttgagtttg atgatctaag
6361  taaaatggtc aatctagatg cgatcattat tgctgttagt cacgaacagt ttaaagatat
```

Fig.4 (Cont II).

```
6421  aacaaagcaa cagtttgata ggctatatgc gcataattct agaaagatta tatttgacat
6481  caaaggtagt ttagataaat ctgagtttga aaaagattat atttattgga gattgtagtg
6541  gcttacgata atgttaaatt tcctcatggt tcgttttttt tggtgactgg aggtgcgggt
6601  tttattggct ctaatttatg tgaagtttta cttagtaagg gttatagagt taggtgttta
6661  gatgatctct caaatggtca ctatcacaat gttgagccgt ttttaactaa ttctaattat
6721  gagtttataa aaggtgatat tagagattta gatacttgca tgaaagcttg tgaaggtatt
6781  gattatgttc tacatcaagc tgcttgggga agcgtaccaa gaagtattga gatgccatta
6841  gtgtatgaag atataaatgt taaaggtgca ttaaatatgc ttgaagcggc tagacaaaat
6901  aacgttaaaa aatttgtcta tgcttctagt tcatcagtat atggtgatga gccaaattta
6961  cctaaaaaag aaggtagaga aggaaatgtt ttatcaccct atgcatttac aaagaaagct
7021  aatgaagagt gggcgagact atacacaaag ttatatggtc tagatactta tggtctaaga
7081  tattttaatg ttttcggtag aagacaagat cctaatggtg cgtatgcagc agttatacct
7141  aaatttatca aacagttatt aaatgatgaa gcgccaacta taaatggaga tggtaaacag
7201  tcgagagatt ttacatatat agagaatgtt attgaggcaa atcttaaagc atgtttagca
7261  gatagtaagt atgccggaga gtcttttaat atagcttatg gaggtagaga gtatcttata
7321  gatttgtact ataatctttg tgatgccttg ggtaaaaaaa tagagccaaa ttttggtcca
7381  gatagagcgg gtgatattaa gcatagtaat gctgatattt cgaaggctag gaatatgctc
7441  ggatataatc cggaatatga ttttgaatta ggcataaagc atgctgttga gtggtattta
7501  attaattaaa tggtatttta atcaagtgta cataaaaaaa gtgtctttta aaattttata
7561  tttatattta ctagcttttt gtattatttt tagtttagaa tttaaatttg ctatattgaa
7621  tattatagtt tatcttccgg cttgtatttt gggtttttta gctcttaaaa aactatttgt
7681  cggaaatatt gttaagaaac aattagcttt cctttttttc tttttctttt tatcaatgat
7741  ttatttaata atagtccaaa taatcttact tgatgcagca tcattgtttc ctcagttttt
7801  atttaacatt ttgatcgcga taggtttttg taactttatt tttgtttcat atgataataa
7861  tgaaaattat ttttttaata tgtctaaaat aatatttttt gttactttct tacaatctat
7921  ttttgtattt ctttcaaggt attatatatt tttaaatgat tggatattct ttttttttagt
7981  gaaaaaaggg aatattgaga tttcgaatgt tattgaatat aagttaagag tattcggact
8041  tagtaacgct ggaggggatg gtttaggatt ttcaattact ataggattat gtttttctat
8101  attttatttt atcaaatata ttaaaggtaa atctatattt accaaactta tgctgtttgt
8161  acctttaatt cttattgtgt tttctaatat tttcatatct agaacatcac tcttaacttc
8221  ttcacttata ttgttaataa caatatttta tatatatatt aaaaaagaaa aattactgtt
8281  tattataata ttggcgctat tctttttatc aatatggata ttgttcaaat taaatttgaa
8341  tttgagttgg gcttttgaaa atatttactc gtacattcaa tctggcgatt tttcacatgg
8401  aagtctaagt gttttaatca ataaaatgct ttttgtgcca gataaccttt tgacttggat
8461  atttggttgt gaggatgtta gtaatactga tattggttat attaaatatt tatactatta
8521  tgggattata tttagtatgt ttttttatat tcttattatt ttcttgtact ttgaaatgag
8581  aaaatgtttt atattttcag agtatcgatc attatttcta ttgttgttaa tagtatgttt
8641  agtttttcaa gcaaaaataa tttttttgac agtaggatta tttactaaat taaccattat
8701  attatttatt ttttctctta aagaaaacag ctttacaact aggagtgtga tttgaaaagg
8761  tttgtacatt taataataaa ccttaaccaa ggtggtgctg aaacaatgct ttataaactt
8821  tgcaaatcta tggataagtc aatatatcat attacgatta tatcacttat gggtagggga
8881  gtatttgcaa ataagttaga agcttatggt gttaaagttt atacattaaa tttaaataaa
8941  tttaatgtac tatttgtatt gtttaaatat attaagatta tcagaagaat aaagcctgat
9001  gttattcatg cttggatgta tcatgcaaat gtaatttcta tattatgcaa gccttttttat
9061  agaaagacta aatatataaa tagtataaga atgggattgg agaattatga tggtcataag
9121  aatcttacaa agtttatgat aaagttgaat gcaaaatttt ctaagttctc agatttaaca
9181  ttaaataatt caaagaaatc attagaagat catcaaaata taggttttaa aaaccaatgc
9241  tttatagcaa atggttttga taaagatgtt tttaaaccga gctttttaaa gtatgaaaaa
9301  tttcgtttaa ataatgattt agatgataat gttaaaatta taggtatcat agcaagaaat
9361  catgctgata aaaatatttc tcgtttctta caaatagcta atttattgtt aaaaagtaat
9421  cctagtttac ggtttttaat tgctggaaga gagtgttcga aaatagatat aggtagttat
9481  ctagataaca aaagtaatgt aaataagttt tttgtatttg aatctgtgga ttctagtgaa
9541  tacttaccag tattagattt atatttgtct acatcaaaag ttgaaggttt tccaaatata
9601  cttgcagaag ccatgctatg tgaagttcct attgttgctt ctaatgttgg agattgtaaa
9661  gatatactta atggatacgg tgaagttttt gagcttagtc aaggtaataa agaaataata
```

Fig.4 (Cont III).

```
9721  gaaaagatta tgaaagtttt agaaacaacg gtagtcatga aaaagcgcat gagagaatat
9781  ataataaata attttagtat agaagctatt ttggaaaaac acgaaaaact ttatcatgag
9841  ggcagtgtct aatgtgtgga gtagtaggct tttactcatt taataaagaa gaaggttttg
9901  actcaataat taatcaatca ttgctttcta taaagcatag agggtcggat gatagtgggt
9961  attggtgcga caatcaagtt actctggggc atactagatt atcaatacac gatataacta
10021 atgcgggaca tcagccaatg ttatctaata gcggtaatac tgctattgtg tttaatggag
10081 aaatatataa ttacttatcc ataaaaaatc agctattaag tgaatattca aatcttaaat
10141 ttaaaagtaa cagtgatact gaggttttgg tcaatgctat tgaactttgg ggtatagata
10201 aaactttaga aaaatgcata ggaatgtttg cttttggagt ttacagtaga aaaactagtt
10261 gcttaatact agctagagat agatttggcg agaagccatt atattttggt atccaaaatg
10321 gtattttggg ttttgcatca gaattgaagg cacttaagcc attaaaggaa tgtggctgga
10381 ggtttgatat agatagagat gctttagcaa catatatgag gtatgcttat gtaccaacac
10441 catactctat ttataaaaat atatctaaac taaatgtagg tagttacata aaatttgatg
10501 ctaaaggtaa tagtaaagag tataaatatt gggattctaa aaaagtacta gattcagaaa
10561 aatataaaga ttcgtatgat caagcaatcc tagatttaga aattaagctt aaaagtacac
10621 tatcaataca aatgcagtca gatgttcctc taggagcatt tttatccgga ggaattgact
10681 caacaactgt agttgctctt atgcaaagta tgtctaaaga taagataaac acttttagta
10741 taggttttaa tcaaaaagaa tataatgaag ctgagcatgc aagagcagta gcaaaacata
10801 taggtacaaa ccacacagat atgtatgtta cagaaagaga tgctcttgat gtaataccaa
10861 aacttgctgg aatatatgac gagccctttg ctgattcatc acaaatacca acgtatcttg
10921 tgagtaaaat agctaagtcg aaagtaacag ttgcactatc aggtgacgct ggtgatgagc
10981 tctttggcgg ttataataga tacttttag caccaaatat tgctaaaaaa atcaaatttg
11041 ctaagttact taaatatgca ccagatgctt ggataaaaaa agctgagata ttaaattttg
11101 gtaagttcgc tttattagca gataaactac taaaactaaa aagagttctc gaaaaagcaa
11161 aaacaaataa agagctttat gtactacttt gttcacaaat aaatgatact agctttgtgt
11221 taggagcaaa agagtatgat atattaagag ataagaatat ttatgatatt ccacaattat
11281 ctttccaaga gtggatgatg tttgttgatt ctaatacata tatgatagat gatatattgg
11341 ttaaggttga tagagcagct atggctaact ctctagagac aagagtgcca tttttagatc
11401 ataatattta tgaatttgct tattccttac caattgacta taaaatacaa cgaggtaacg
11461 gaaaaagaat tttgaaagat ttgttatata aatatgtgcc agaaagtttg gtcaataggt
11521 ctaagatggg gtttggtatt ccgcttgcta aatggttaag agaagattta cgagagtggg
11581 cagataattt actggattat agtaaaatag acaagcaagg ttacttaagt cctgaggtgg
11641 tgcaaaaata ttggcaagag catttgagtg gtaaaagaaa ttggcaagca atattatgga
11701 atattctaat ttttcaggag tggttagata atgagtaaag taaatgtaac aaaaccatac
11761 ttaccagata taaataaata taaaagctat gtaaataaaa tatacaaaaa tggatggctt
11821 actaataatg gtccgttagt gcaagagcta gaaaaaagac ttgcaaagta tctaggtgtt
11881 aaaaatatag ttttagtatc aaatggtaca attgcattag aaatcgcgta tagagcgtta
11941 ggagtcaaag gaagtgcaat tactactcca ttttcatttg ttgctactac atcttcattg
12001 gtttctaaca atgtaaaacc agtgtttgtt gatattgatg agaatactct aagtatagac
12061 gtctctaaaa ttaagtatgc tattgaagag gatacttcag ctattgtgcc agttcatgtg
12121 tttggaaatg gttgtgaagt tgaaaaaata gacatgctgg ctaaaaaaca taacttaaaa
12181 gttatttatg atgcagcaca tgcttttgat gttaagtata agggtgagag tatattaaac
12241 tatggtgata tttcgacatt aagttttcat gcaacaaaga tttttcattc tattgaagga
12301 ggtgcgctta tcattaatga tgatagtctt gttgaaaaag ttcgttattt cattaatttt
12361 ggtatagaaa gctcagaatc aataccttac ttaggtacta atgctaaaat gaatgaattt
12421 gaggcggcta tgggactttg tgttctagat gatattatag aaattaagag caaaaggaaa
12481 gttattacag agatatatga ggctgggtta gatggattgg taaagtttca agaacagaat
12541 cagcattcta gtaggaatta tagctatttt ccagtaatat ttaggactga ggaggaactt
12601 ctcagagtac agaaagcact aatacaaaat gatataatat cgcgtagata tttttatcca
12661 tcattagata gtcttagtta tatagagcca aagcagtata tgccaatctc aagagatata
12721 tctaaaagaa tattatgttt gccaatttat gcagagttag aagacgataa aattaataaa
12781 ataattaata atatcaaaga ggtttcctca tgaaaaaaat atttgttgtt acagataata
12841 gaactattct aagtgatttt aaaaatatca ttggtagtaa aaatgatgta caggttgatt
12901 atttttgtag tttcaagagt caaacttctt ttgccaaaga aatatataac agtgagatta
12961 agccaataga tatgaaaaaa aatggcaatg atcttattgg taagtatgat ttaggttttt
```

Fig.4 (Cont IV).

13021 cttgtcattc gaaacaatta tttccagcaa aattagttaa ttcagtatta tgtataaata
13081 ttcatcctgg acttaatcca tataatagag ggtggtttcc acaggtcttc tctattataa
13141 ataaactacc tataggagca actattcatg tgatggatga agagatagat catggagata
13201 taatcattca ggaagaagtt gaagttaatt ctttcgaaaa ctcttttgat gtttatgcta
13261 aagttcaaaa aaaagaagtt gagttgttca ctaaagtcat agatgatatt ttgaataata
13321 agttcactcg aatcaaacct aactccgaag gcaactataa ttcaattcat gattataaaa
13381 acatgtgtga aattgattta gataaaatag taacaatgcg ggaagcaatt gactatctaa
13441 gggctatgac acaccctcca tataaaaata gttatttcat tgatgagcat ggaaataaag
13501 tatttgttgc tcttgaactt gaaaagataa gttagaaaaa tgagccttaa aaaaaataca
13561 atatcaaatt atataacaca actatatact agcttaattg gtattgttat acttcctttg
13621 tatttacaac atttaagtca tgatgcattt ggtctgattg gttttttttac agtttttcaa
13681 acgtggttac ggttgttgga tgttggtata acaccaactt tatcaagaga agtggctcat
13741 gttagaggta gtactgatga ctatcattac ttacgcaagt tggttagatc gttagagcta
13801 tttttcatta ttgttggtgt tctggtattt attgtaatta gtacacattc aaggtatata
13861 tccacctctt ggttacatat aggctcgcta gatgctgata gtgtaagtgt atgtattgca
13921 cttatgggtt taatgtttgc attaagatgg gtgtctgatc tatatggtgg tggtttgcgt
13981 ggctttgaaa gacaggttct ttataataat ttaagtatca tacaaacgac actacagttt
14041 attggtggat tattatttat ctgctatgtg tctactaata ttatgtatta ttttgtatat
14101 cagacaataa ttgcgatact atatctagta tgtattgcaa ttgcatttta taaaatacta
14161 ccatcatcat ttagcgtggg tttaaggttt gattttaaaa taattagaaa agtgcttcca
14221 tttgcactag gcattgcata ttctacaaca gtttggatta ttgtcactca atctgataaa
14281 ttagtgttct cacatgtatt accattatct gagtatggtt atttatcttt attgatagtg
14341 atatctagtg ctgttacgat attgtcctct ccgattagca tagctattca gcctagaatg
14401 acaatgctat tagcccaaca aaatgtaaaa ggaatggaaa gcttatattt aaaatcatcc
14461 ttgatctcaa ttacttttt atctgctgta gtaacatgtg ttttgatgta ttctcatcag
14521 ctgttgcagt catggacagg aagtatggaa attgctaatt ggggtagtaa tatcttaaat
14581 atatatgttt tatcagcatc tattatttgt ataatatcat ttcaatattt tttacagtat
14641 gcttatggta agttaaagct acataataca tataatacaa ttagtttagt atttttgct
14701 cctatagtta tatatactgc ttataattat ggagtgtata ctacagcact attatggctt
14761 ggatatgcta tagtggggct gataatctgg atgcctattg tacaccatgt atttgctaaa
14821 ggtatcaata ggtatttttt tataaattta gcagttatta ctatagtatg ttttttatta
14881 tcgttaatat ttaagggttg gtatatttat ccaagtaaaa ttgggttggt agaattaata
14941 ttgattgggt ttgcattttt atttatacaa atttgtatag agtatgtttt gtttcggtac
15001 aaggttttga ggtgtataga tgattaaagt ttcagtatgt gtgatgacat acaatcaaga
15061 aaagtatatt ggtcaatgtt tagagtcttt ggttactcaa gagactgatt ttgactttga
15121 gataatcgtt ggagatgatt tttctacaga tggtacaaga gatgttattc aagagtatca
15181 aaaaaagtat ccggatatca taaagccagt ttttagagat aagaatgtgg gaattactga
15241 aaatattaaa gaaatctatt ttgttgcaaa tggtgagtat atagctcata tggatggtga
15301 tgattatgca ttgcctggta aacttcaaat tcaggctgat ttttttggata ataatccaag
15361 atgtacggga gtttttcata atataaatat actctatcca aatggtaata tacaacatag
15421 taggtttgct tgttcaaata agagtatatt caatttatca gacactttac gcggagttgc
15481 tgttggtgca aatagttcaa aaatgttcag aacatcggtt ttggatgatt tgattttacc
15541 ggatatagag cttctagatt attattttca tgttataaca gcagaaaaag gttatttaag
15601 ttttttaaat tctaatgaat cctatagtgt gtacagaaaa ggtattggta tcacatctaa
15661 gtctaaggaa aaaatctata atacttatgc tggattattt gaatatttttt tggatagata
15721 tcctgaagag aaattaaata tttgtatccc tgttgtgcaa atgataattt cggctattaa
15781 agggagatgt tttattagtg ctattcgtct attcaaaatt ttaattagat caagatgtat
15841 tccattagta agttggttta aatatagatt tgaaaaataa atatcattta gaggattatg
15901 tgaaatgaag ggaataattc tagctggtgg cagtggtaca aggctatatc cacttacctt
15961 gggtgttagc aaacagctgc tacctgttta tgacaagcca ttgttatact atccactatc
16021 tgtgcttatg cttgcaggta ttagggagat attaattatc tctacagtgc gtgatatctc
16081 acttatccaa gagcttcttg gtgatggttc acaatttggt atacagttga gttataaaat
16141 ccagccatca ccagatgggc ttgctcaagc atttattctt ggtgaggagt ttttggcggg
16201 tgactcagct tgtttgatat taggagataa tatctactat ggtcaaggta tgactacaat
16261 gctagagtct gcaagagcac agtgtggagg tccagctggt ggcgcttgtg tttttggtta

Fig.4 (Cont V).

16321 ttatgttaat gatccgcata gatatggtat agtcgaattt gataagcaaa aaaatgtaat
16381 ttcggtagag gaaaagccac agaatcctaa gtcacactat gctatcacag gtttatattt
16441 ttatgataat aatgttgttg agtatgctaa acaagtcaaa ccatctgcac gtggtgagct
16501 agagattact tcacttaatg agttatatct aaaagaaaat aagctaaatg tcgaactctt
16561 agggcgtggc tttgcttggc ttgatgctgg tacgcatgat tcattgctag aggcaggtca
16621 atatgtcgca actattgaga aaagacaagg gcttaaaatt gcatgtttgg aagaaattgc
16681 atggcgtaaa ggctttatct caacacaaca agttctagct caagctgaaa aactttctaa
16741 gacagagtat ggtcagtatc tgaagaattt aattaaggat ggtttataaa ttaatccgtc
16801 atacccatga aggtgggtat ctcataaaag ttggatatgt tttggagatt ccaatctgcg
16861 cagtaatgac aggtttggta atatatagcg atgttttaca atgactaaaa atggttttat
16921 gtatattctt acaaataagg ataatactgt tctgtacata gttgtaacat ctaatttgat
16981 aaaaagaatg tatgagcata aacatagcct tgcagatggt tttactaaaa atataatgtt
17041 aataagttag tttattttga aatttatgaa gatataaaag cagcaattct gtgagaaaag
17101 cagttgaaaa aatgaaacag atcttggaaa gaacgaatta ttaatgagat gaatccgaat
17161 tggaatgatt tatatgaatt aatatgtgag taaaactttt gtcttactgg tgcagatagg
17221 tatctctaaa tatcagatgt gattgggaga ttaccgccta cgcggtaatg acaagtttat
17281 gcggtaatga tagtttagtg agagaatgac tagtcactat aggaatgatg atgtaatgag
17341 gaatgaaaaa atgaactaca aaccaaaaaa tatcctagta acaggtgcgg cgggatttat
17401 tggtagtaac tatgtgcgta tgatgttatc acgctatagt gatatcaaaa taatctcgta
17461 tgataagctt acttatgcgg gtagtttaga taatctaaaa gacttgaata atgaacataa
17521 ccatactttt ataaaaggtg atatttgtga tgaagtttta gtatatcaaa cactgaaaga
17581 atataaaatt gatacgatag tacattttgc tgcagaatcg catgttgata attcaattgc
17641 taatccaaag gtatttttag aaacgaatgt gataggtaca tttacacttt tagattgtgc
17701 taaaaggtat tggttagatg agctaggttt agaagaaact agttgtaggt ttcatcatgt
17761 atctactgat gaggtatatg gtaccttggc aaaagatgaa ccagccttta ctgagattaa
17821 ggcttatgag ccaaattcac cgtattcggc atctaaggcg ggatctgatc atatttctag
17881 agcatatcat catacctata aacttccggt aacaatttca aattgttcaa acaactatgg
17941 accataccaa catcgagaga aattaatccc tgtagtgata aatagttgta taaactacaa
18001 gcctattcct gtttacggag atggttcgaa tattcgagat tggctatatg tagaagatca
18061 ctgcgatgct atccagacaa ttgttgagaa aggagtggtt ggagaggttt ataatattgg
18121 tggtattaat gaagttgata atctaacctt ggtaaaaact atctgtaaac taatggatga
18181 atataaacca gaaaatgctc cacattctaa cttaatcaca tttgtggaag atagaaaagg
18241 acatgattgg cgttatgcta ttgataacag caagattcag aatgagttag gatggaagcc
18301 atcacaagat tttgataaga tgtttagaca aactattgag ttttatctat agcttaaata
18361 tttatcttat gagtatctct aaaaaatcaa tttaatttat ttttgtgtta aaaagtagtg
18421 ttcgcaagaa tatagttaat ccgaaagata tttgtagaaa aagatatttg tagaaatgtt
18481 ataatgtcta ataaaaatgc catcatatag ccaagatttt agagacatcg taattaataa
18541 acatgaagaa ggtatgacgg agttcgagct gagtaagttt tttaacatag ataagcgtac
18601 agttgtttca tggatagagt tttataaaag aaccggagat tatagttcaa agcaaggagt
18661 tggttgtggc agagtcgcta gctttaccga taaaacattg attgaacagt atttgataga
18721 tcatccagat gcaagtgcat tagatataaa agaagcatta gcccctgata ttccaagaag
18781 tacattttat gattgtctta atagacttgg ttttagtttt aaaaaaagac tccaaaatat
18841 aagcaaagaa aagaacatga aaggttggag tatatagaaa aactaaaaga aatagccaat
18901 aaatttgatg tacaaatatt atatctacct ccgtactctc cagatttaaa tcctattgaa
18961 aaggtttggg ctaactatta aaaaaatatt tagaaaagtg aataatagtt ttgaaaaatt
19021 ttgtgatgct atctcttatg tgtttaacaa aatactctcg gattaactat atcatgctgc
19081 taaaatattc ttggtattct ctggtcaaaa ctgacataat gatgctctac tttgtataag
19141 gtttgctaca aatattatct aaacaaacat acaaaggtaa tttttagaga tcctattata
19201 aacctactat ctaaatttag taagttaagt tatgacaata tttaatttgc tgatttattg
19261 ttgaatatat tagctttcta tataattaat caatatcaaa gttatttagg tttttataat
19321 gattactcct attatcttat ctggaggatt cggctcaagg ctatggccac tatcacgaga
19381 ggcatcgcca aagcagttta tcggcttggt tgatgaacat agtctattag aaaatacaat
19441 taagcgacta gataatgtca aggatataac ttcacctgta gttgtctgta atgaaagtca
19501 tagattccaa gttgctgaag tgttgcggaa aatcaataaa aaaggcgata tactcctaga
19561 gccattagcc agaaatactg ctccagcaat tgcacttgca gcactacatt tagctattaa

Fig.4 (Cont VI).

```
19621 tgatccaaat acaattatgc tagttttagc tgctgaccat catattgaaa atctggagat
19681 ttttcatcaa gctatcgaaa aagcacagca aaaagttatt aaagatgatt ctttagttac
19741 ctttggcatt acaccaactt gtcctcatga aggctatggt tatattaaac aaggggtaca
19801 gactactgta aatggagttt ataaggtaga taaatttgtt gagaagccta gtgtggtcgt
19861 tgcacaagag tatttagata gtggcaaata ctattggaat agcggtatgt ttatgttcac
19921 agctagagtg tatttagagg ttttagagaa gttacagcca gagatttaca gaggatgtga
19981 aaaaacttat caaaagtcac agcaggattt agattttgtg cgttttgata aacaaagctt
20041 tgccctagtt caatcacagt caatagacta cgcagttatg gagaaagcaa ctaatgttgc
20101 tatagtgcct atgcaacaaa gtggctggtc tgatgttggc tcttgggact ctttgtatga
20161 tattgctgca aaagatagtt gtggtaatgt ggttattggc gatgtgatta ctagtaatgt
20221 caaaaatagt tatttacgct cgcatgatcg tttattggct gcagtcggag ttaatgattt
20281 aataattgtt gaaacagcag atgctatact tgtcgcggat aagaacaaaa ctcaagatgt
20341 caaaaaaata gtcgaagttt tgaaaattca gcagcgaagt gaattattac agcataagca
20401 aatttataaa ccttggggtt cagcgacaat attagaggat aagtctggtt ataagataca
20461 ggcgattcaa cttgaaccgg gcaagaagtt atcattacag caacattatc accgtagtga
20521 gcattggatt gtgatttctg gaactgctac ggtaactatt ggtactacta agtctattgt
20581 tagaccaaat gagtctgtat atataaaaat aggcgaatct cacagacttg aaaataatgg
20641 caagattcca gttattctta tagaagtaca agttggagaa tatataagtg aagacgatat
20701 tgttagacta gatacaagta gttaatataa aaacaattag atagaaaaaa atataatgag
20761 acaaactata ataaaagaaa taatcaaatc tagcggcgta aagtttggta ctagtggagt
20821 tagaggtctt gtttcagcta tgacagataa gatctgttgg ctttatacaa aagcttttat
20881 tcaattccta gagcaaaaat actctattgc taagggtact aaaattgcta tagctcatga
20941 tctacgtgag agtagcccta gaataacaac agttgttatt aaagctatca tagatagtgg
21001 tcatgagcca atatactgtg gtgagatacc atcaccagct gtaatgctat atggtatatc
21061 taatcagata ccgtcagtta tggttactgg tagtcatatt ccagaggata gaaatggtat
21121 taagtttaat actccatatg gtgaagttct caaagaagat gaagaaatga ttgttagcca
21181 aactatcagc attgatgaaa gtatttttga taaaaatggc atgtttttac aaaaactaga
21241 attaccagag cctagtaagc aagcatatac acagtatatt gacaggtatg tagatttttt
21301 ccctaacaac tgtctagcag gtaagactat agggctttat cagcactcat ctgtaggacg
21361 agagatagtc aaagagattc tagagaaact aggtgctaag gttatcttgc tagaattttc
21421 cgaaaaattt gtatctgtag ataccgaggc aattcgccag gaagatgtaa agcttgctaa
21481 gcagtgggca agcaagtata aagttgatag tatagtttca actgatggcg atgctgatag
21541 gccactagtt agtgatgagt atggcaattg gctaaaaggt gatattttag gtgtactgac
21601 agctaaatat ctccaagcca atgttatcgt gacaccagta agtagcaata ctgtggcaga
21661 aaagataggt tattttagta acgtgattag aactaaaata ggctcgccgt atgtaattgc
21721 tgcaatgaat gaattactct caaataatca aaatgctgtg gttggatatg aggcaaatgg
21781 aggatttcta ttggctagtg atatttgtaa agatgataaa actctaaaag cgctgcctac
21841 aagagatgct gttataccaa tgttggctgt aatgatgcta tctatcaact ctaataaaac
21901 cgtgtcagag cttttatttg atttgccatc tcgatataca gcaagtagta aaattgatga
21961 ttttgcttcc gagaaaagcc aagaaatctt gaagtcaata ttagcaggtg aatcagatct
22021 tttagataaa attatatcgg agcattttga tggtaaaaat agcattgaaa atatcgatac
22081 tacagatggt gttagagtaa ctttgacaaa tcaagatatt atccatctta gaccatctgg
22141 taatgctcca gagcttaggt gctatacaga ggcagctagt gatgagcagg caaaaagttt
22201 aaatcaatat tgtgtggatt tgattaacaa aaacatttga agatcagtca aaaatattcc
22261 ctaacttttc tcttcaccat tgaaccatta ctaaccttat ctatagctag ccacagataa
22321 aaatgtcatg ctggatttat ttcagcgttt cattataaat atcaatttta ttgagatcct
22381 gaaactagtt caggatgaca g
```

Fig.5.

SEQ ID NO 2
MNYHIKEVFWSIILSFLKSQKGIHTNDEAKLRLFIEAVFYVLRT
GCQWRMLPFYYGKYRSIHKRFKDWCDKDIFSRLFKSVQNPDLQEVMLDSTIARAHACA
TGYDKDDNQAIGRSVGRITTKIHAMTDALGNPIEILLSEDKTHDSKVAIDLLKNVYNT
KVIADRAYHSNEIRQHIQGISSEAVIPCKSNTLNHIPFDSHVYKERHLIENFFSKIKH
FRRVFSRFDKTILAYIGMIKLACTFIWLR

SEQ ID NO 3
MSFYDNRTLNFVVIIVLTIITVNWTFYIFKQDVNLHFLLALVLL
RCLSSFLLLRDYMASWRKSTQKTFLRKAFINLPVFFIVALFFYGKVTFSLIFSEFLFY
VFLISLSVYFYWYLMNRGSVDKSKTAVIYGAGAAGTKIAQELASAGYRIKCFVDDNET
LQKRSIDSKKVLSKAELTKLLLSSRFDLLVIALPRNANQVVKNIYKEFEKDFNQIRIM
PPLEEILQDENFMSQLKPVSLYDLLARDTKSLDKESISNFIKNKVVLVTGAGGSIGSE
IVHQCIKYQAKELILVDHSEFNLYKITEECSHFNINSVLCSVCDRKALAEVFQKYTPN
IVFHAAAYKHVPLVEENISRAIRNNILGTKNAIDLAIEAGVESFILISTDKAVRPTNV
MGATKRVCELYLQNVDPKNTKLAAVRFGNVLGSSGSVIPKFEEQIRKGGPVTVTHPEI
TRYFMLIPEACELVLQAGAIAKNSEVFVLDMGQPVKIIDLAKQFIRLSGRGDIDIKIV
GLRPGEKLYEELLIEEDDVSTDYKDIFIGRRTFYDINTLNQDIESLIKDDVDQLVILK
KIVPEFEHRLNG

SEQ ID NO 4
MFYEVFKRLLDILLSFMGLLLLSPIFLIIIFMIKKDSKGPIFFK
QKRYGKDKQFFYIYKFRTMYVDTPKDMPTHMLQDPSKCITKVGGFLRKSSLDELPQII
NILKGEMSIVGPRPALWNQDDLIAQRDKYGANAVPVGLTGWAQINGRDELPIPDKAKL
DGDYVKNKSTWFDLKCIFLTVFSVFAKKGVVEGGTGALGNKEDLK

SEQ ID NO 5
MKKRILVTGLSSYIGNSFAAKYNSDFSIDKISLRDVSWANIDLS
GYDAVLHVAGIAHTSKDPKLKEKYYKINTQLTYDLAKQAKDQGVRQFVFLSSIIVYGD
SAPIGQQKVITKYTEPKPDDFYGDSKLQTEIKLNSLASDDFNIAIIRPPMVYGEGSKG
NYPKLVKLAKYTFIFPNINNQRSVISIDNLSKEIAEIILQTKHGVFLLQDNEYFCTSQ
FIKNYRKDVLGKRTYLTKIFNPIIRLLAKKVDFINKVFGNLTYEK

SEQ ID NO 6
MRSKLLFIANDFDIVIYRFRREVIESFAAKEYEIVLVTPYSKKA
EVFCKSLGVKYINVDIDRRGKNPFKDLLLLFNYFKIIKKEKPDYIFSYTIKPNLYVGL
VNLFFRKKFYPNVTGLGSVFANHGIVQKFIISLYKLSFKSTTKVFFQNEQNKKLFIAK
KIISGEKSILLPGSGVNLDENKYVDYPKDQGILKFVFLGRIMKEKGIYELLEAFAILE
KKYKNISLDIYGFCDENKSNFMGKVNTIKSVKFYGFTDNTKEKIASAHAVVLPSYHEG
MSNVLLEAAAIGRPVIASDIPGCREIFDDGLSGLSCNPNDVSSLRNSLEQFINMSYTD
KIAMSYKARAKIEKDFDRSIVVNAYLQQN

SEQ ID NO 7
MSLYEDIVAKREKVSLVGLGYVGLPIAIAFAKKIDVLGFDICET
KVQHYKDGFDPTKEVGDEAVRNTTMKFSCDETSLKECKFHIVAVPTPVKADKTPDLTP
IIKASETVGRNLVKGAYVVFESTVYPGVTEDVCVPILEKESGLRSGEDFKVGYSPERI
NPGDKVHRLETIIKVVSGMDEESLDTIAKVYELVVDAGVYRASSIKVAEAAKVIENSQ
RDVNIAFVNELSIIFNQMGIDTLEVLAAAATKWNFLNFKPGLVGGHCIGVDPYYLTYK
AAELGYHSQVILSGRRINDSMGKFVVENLVKKLISADIPVKRARVAIFGFTFKEDCPD
TRNTRVIDMVKELNEYGIEPYIIDPVADKEEAKHEYGLEFDDLSKMVNLDAIIIAVSH
EQFKDITKQQFDRLYAHNSRKIIFDIKGSLDKSEFEKDYIYWRL

Fig.5 (Cont I).

SEQ ID NO 8
VAYDNVKFPHGSFFLVTGGAGFIGSNLCEVLLSKGYRVRCLDDL
SNGHYHNVEPFLTNSNYEFIKGDIRDLDTCMKACEGIDYVLHQAAWGSVPRSIEMPLV
YEDINVKGALNMLEAARQNNVKKFVYASSSSVYGDEPNLPKKEGREGNVLSPYAFTKK
ANEEWARLYTKLYGLDTYGLRYFNVFGRRQDPNGAYAAVIPKFIKQLLNDEAPTINGD
GKQSRDFTYIENVIEANLKACLADSKYAGESFNIAYGGREYLIDLYYNLCDALGKKIE
PNFGPDRAGDIKHSNADISKARNMLGYNPEYDFELGIKHAVEWYLIN

SEQ ID NO 9
VYIKKVSFKILYLYLLAFCIIFSLEFKFAILNIIVYLPACILGF
LALKKLFVGNIVKKQLAFLFFFFFLSMIYLIIVQIILLDAASLFPQFLFNILIAIGFC
NFIFVSYDNNENYFFNMSKIIFFVTFLQSIFVFLSRYYIFLNDWIFFFLVKKGNIEIS
NVIEYKLRVFGLSNAGGDGLGFSITIGLCFSIFYFIKYIKGKSIFTKLMLFVPLILIV
FSNIFISRTSLLTSSLILLITIFYIYIKKEKLLFIIILALFFLSIWILFKLNLNLSWA
FENIYSYIQSGDFSHGSLSVLINKMLFVPDNLLTWIFGCEDVSNTDIGYIKYLYYYGI
IFSMFFYILIIFLYFEMRKCFIFSEYRSLFLLLLIVCLVFQAKIIFLTVGLFTKLTII
LFIFSLKENSFTTRSVI

SEQ ID NO 10
LKRFVHLIINLNQGGAETMLYKLCKSMDKSIYHITIISLMGRGV
FANKLEAYGVKVYTLNLNKFNVLFVLFKYIKIIRRIKPDVIHAWMYHANVISILCKPF
YRKTKYINSIRMGLENYDGHKNLTKFMIKLNAKFSKFSDLTLNNSKKSLEDHQNIGFK
NQCFIANGFDKDVFKPSFLKYEKFRLNNDLDDNVKIIGIIARNHADKNISRFLQIANL
LLKSNPSLRFLIAGRECSKIDIGSYLDNKSNVNKFFVFESVDSSEYLPVLDLYLSTSK
VEGFPNILAEAMLCEVPIVASNVGDCKDILNGYGEVFELSQGNKEIIEKIMKVLETTV
VMKKRMREYIINNFSIEAILEKHEKLYHEGSV

SEQ ID NO 11
MCGVVGFYSFNKEEGFDSIINQSLLSIKHRGSDDSGYWCDNQVT
LGHTRLSIHDITNAGHQPMLSNSGNTAIVFNGEIYNYLSIKNQLLSEYSNLKFKSNSD
TEVLVNAIELWGIDKTLEKCIGMFAFGVYSRKTSCLILARDRFGEKPLYFGIQNGILG
FASELKALKPLKECGWRFDIDRDALATYMRYAYVPTPYSIYKNISKLNVGSYIKFDAK
GNSKEYKYWDSKKVLDSEKYKDSYDQAILDLEIKLKSTLSIQMQSDVPLGAFLSGGID
STTVVALMQSMSKDKINTFSIGFNQKEYNEAEHARAVAKHIGTNHTDMYVTERDALDV
IPKLAGIYDEPFADSSQIPTYLVSKIAKSKVTVALSGDAGDELFGGYNRYFLAPNIAK
KIKFAKLLKYAPDAWIKKAEILNFGKFALLADKLLKLKRVLEKAKTNKELYVLLCSQI
NDTSFVLGAKEYDILRDKNIYDIPQLSFQEWMMFVDSNTYMIDDILVKVDRAAMANSL
ETRVPFLDHNIYEFAYSLPIDYKIQRGNGKRILKDLLYKYVPESLVNRSKMGFGIPLA
KWLREDLREWADNLLDYSKIDKQGYLSPEVVQKYWQEHLSGKRNWQAILWNILIFQEW
LDNE

SEQ ID NO 12
MSKVNVTKPYLPDINKYKSYVNKIYKNGWLTNNGPLVQELEKRL
AKYLGVKNIVLVSNGTIALEIAYRALGVKGSAITTPFSFVATTSSLVSNNVKPVFVDI
DENTLSIDVSKIKYAIEEDTSAIVPVHVFGNGCEVEKIDMLAKKHNLKVIYDAAHAFD
VKYKGESILNYGDISTLSFHATKIFHSIEGGALIINDDSLVEKVRYFINFGIESSESI
PYLGTNAKMNEFEAAMGLCVLDDIIEIKSKRKVITEIYEAGLDGLVKFQEQNQHSSRN
YSYFPVIFRTEEELLRVQKALIQNDIISRRYFYPSLDSLSYIEPKQYMPISRDISKRI
LCLPIYAELEDDKINKIINNIKEVSS

SEQ ID NO 13
MKKIFVVTDNRTILSDFKNIIGSKNDVQVDYFCSFKSQTSFAKE
IYNSEIKPIDMKKNGNDLIGKYDLGFSCHSKQLFPAKLVNSVLCINIHPGLNPYNRGW
FPQVFSIINKLPIGATIHVMDEEIDHGDIIIQEEVEVNSFENSFDVYAKVQKKEVELF
TKVIDDILNNKFTRIKPNSEGNYNSIHDYKNMCEIDLDKIVTMREAIDYLRAMTHPPY
KNSYFIDEHGNKVFVALELEKIS

Fig.5 (Cont II).

SEQ ID NO 14
MSLKKNTISNYITQLYTSLIGIVILPLYLQHLSHDAFGLIGFFT
VFQTWLRLLDVGITPTLSREVAHVRGSTDDYHYLRKLVRSLELFFIIVGVLVFIVIST
HSRYISTSWLHIGSLDADSVSVCIALMGLMFALRWVSDLYGGGLRGFERQVLYNNLSI
IQTTLQFIGGLLFICYVSTNIMYYFVYQTIIAILYLVCIAIAFYKILPSSFSVGLRFD
FKIIRKVLPFALGIAYSTTVWIIVTQSDKLVFSHVLPLSEYGYLSLLIVISSAVTILS
SPISIAIQPRMTMLLAQQNVKGMESLYLKSSLISITFLSAVVTCVLMYSHQLLQSWTG
SMEIANWGSNILNIYVLSASIICIISFQYFLQYAYGKLKLHNTYNTISLVFFAPIVIY
TAYNYGVYTTALLWLGYAIVGLIIWMPIVHHVFAKGINRYFFINLAVITIVCFLLSLI
FKGWYIYPSKIGLVELILIGFAFLFIQICIEYVLFRYKVLRCIDD

SEQ ID NO 15
MIKVSVCVMTYNQEKYIGQCLESLVTQETDFDFEIIVGDDFSTD
GTRDVIQEYQKKYPDIIKPVFRDKNVGITENIKEIYFVANGEYIAHMDGDDYALPGKL
QIQADFLDNNPRCTGVFHNINILYPNGNIQHSRFACSNKSIFNLSDTLRGVAVGANSS
KMFRTSVLDDLILPDIELLDYYFHVITAEKGYLSFLNSNESYSVYRKGIGITSKSKEK
IYNTYAGLFEYFLDRYPEEKLNICIPVVQMIISAIKGRCFISAIRLFKILIRSRCIPL
VSWFKYRFEK

SEQ ID NO 16
MKGIILAGGSGTRLYPLTLGVSKQLLPVYDKPLLYYPLSVLMLA
GIREILIISTVRDISLIQELLGDGSQFGIQLSYKIQPSPDGLAQAFILGEEFLAGDSA
CLILGDNIYYGQGMTTMLESARAQCGGPAGGACVFGYYVNDPHRYGIVEFDKQKNVIS
VEEKPQNPKSHYAITGLYFYDNNVVEYAKQVKPSARGELEITSLNELYLKENKLNVEL
LGRGFAWLDAGTHDSLLEAGQYVATIEKRQGLKIACLEEIAWRKGFISTQQVLAQAEK
LSKTEYGQYLKNLIKDGL

SEQ ID NO 17
MTSHYRNDDVMRNEKMNYKPKNILVTGAAGFIGSNYVRMMLSRY
SDIKIISYDKLTYAGSLDNLKDLNNEHNHTFIKGDICDEVLVYQTLKEYKIDTIVHFA
AESHVDNSIANPKVFLETNVIGTFTLLDCAKRYWLDELGLEETSCRFHHVSTDEVYGT
LAKDEPAFTEIKAYEPNSPYSASKAGSDHISRAYHHTYKLPVTISNCSNNYGPYQHRE
KLIPVVINSCINYKPIPVYGDGSNIRDWLYVEDHCDAIQTIVEKGVVGEVYNIGGINE
VDNLTLVKTICKLMDEYKPENAPHSNLITFVEDRKGHDWRYAIDNSKIQNELGWKPSQ
DFDKMFRQTIEFYL

SEQ ID NO 18
MPSYSQDFRDIVINKHEEGMTEFELSKFFNIDKRTVVSWIEFYK
RTGDYSSKQGVGCGRVASFTDKTLIEQYLIDHPDASALDIKEALAPDIPRSTFYDCLN
RLGFSFKKRLQNISKEKNMKGWSI

SEQ ID NO 19
MITPIILSGGFGSRLWPLSREASPKQFIGLVDEHSLLENTIKRL
DNVKDITSPVVVCNESHRFQVAEVLRKINKKGDILLEPLARNTAPAIALAALHLAIND
PNTIMLVLAADHHIENLEIFHQAIEKAQQKVIKDDSLVTFGITPTCPHEGYGYIKQGV
QTTVNGVYKVDKFVEKPSVVVAQEYLDSGKYYWNSGMFMFTARVYLEVLEKLQPEIYR
GCEKTYQKSQQDLDFVRFDKQSFALVQSQSIDYAVMEKATNVAIVPMQQSGWSDVGSW
DSLYDIAAKDSCGNVVIGDVITSNVKNSYLRSHDRLLAAVGVNDLIIVETADAILVAD
KNKTQDVKKIVEVLKIQQRSELLQHKQIYKPWGSATILEDKSGYKIQAIQLEPGKKLS
LQQHYHRSEHWIVISGTATVTIGTTKSIVRPNESVYIKIGESHRLENNGKIPVILIEV
QVGEYISEDDIVRLDTSS

Fig.5 (Cont III).

SEQ ID NO 20 MRQTIIKEIIKSSGVKFGTSGVRGLVSAMTDKICWLYTKAFIQF
LEQKYSIAKGTKIAIAHDLRESSPRITTVVIKAIIDSGHEPIYCGEIPSPAVMLYGIS
NQIPSVMVTGSHIPEDRNGIKFNTPYGEVLKEDEEMIVSQTISIDESIFDKNGMFLQK
LELPEPSKQAYTQYIDRYVDFFPNNCLAGKTIGLYQHSSVGREIVKEILEKLGAKVIL
LEFSEKFVSVDTEAIRQEDVKLAKQWASKYKVDSIVSTDGDADRPLVSDEYGNWLKGD
ILGVLTAKYLQANVIVTPVSSNTVAEKIGYFSNVIRTKIGSPYVIAAMNELLSNNQNA
VVGYEANGGFLLASDICKDDKTLKALPTRDAVIPMLAVMMLSINSNKTVSELLFDLPS
RYTASSKIDDFASEKSQEILKSILAGESDLLDKIISEHFDGKNSIENIDTTDGVRVTL
TNQDIIHLRPSGNAPELRCYTEAASDEQAKSLNQYCVDLINKNI

Fig.6.

SEQ ID NO 1

atatttatttttgtgcacagaacctaatttgcatttttgtgcacaaagaaaatttttttg
atataatagactttaataggatattttctaaaaattaacaaatgtctttctacgataata
gaacgcttaatttcgtggtaataatagttttaactattattactgttaattggactttct
atattttcaagcaagatgttaatttacatttttacttgcattagttttgctgagatgct
tgtcatctttttactacttagagattatatggctagttggcgtaagtcgactcaaaaaa
ctttttacgtaaggcttttattaatttgccagtatttttcatagtggcattatttttttt
atggcaaagtcacttttttcgttgatattctctgagtttttatttttatgtttttttgatca
gtttaagtgtctacttttattggtatttgatgaacagaggatcagtggataaaagtaaaa
ctgcggttatttatggtgcaggtgctgcaggaacaaagattgctcaagaacttgcttctg
ctggttatcgcatcaaatgttttgttgatgacaatgaaactttacaaaaaagaagtattg
atagtaaaaaggttctatctaaagctgaattaacaaaactattgctatctagtagatttg
acctttttggttattgcattgccaagaaatgcaaaccaagtagtcaaaaatatatataaag
aatttgaaaaggattttaatcagattagaattatgccgcctcttgaggaaattcttcaag
atgagaattttatgtcacagttgaagcctgtttcactctatgatctattagcgcgtgata
ctaagagtttagataaagaatctatctctaattttatcaaaaataaggtggtgctagtca
caggagctggaggtagtataggttctgaaatagtacatcaatgtatcaagtatcaggcaa
aagagttgatattggttgatcatagtgagtttaacttatataaaattactgaggagtgta
gtcattttaatatcaatagtgtgctatgttctgtttgtgatagaaaagcattggctgagg
tttttcaaaagtatactccaaatatagtatttcatgctgctgcctacaagcatgttccct
tagttgaggagaatatctctagagcaattagaaataatatcttaggtactaagaatgcta
tagatctggctatagaagctggtgttgagtcatttatattgatttccactgataaagcag
tgcgaccaacgaatgttatgggggctaccaagagagtttgtgagctgtatttacagaatg
ttgatcccaaaaataccaagcttgctgcagtgcgttttggtaatgtgcttggtagtagtg
gcagtgtgattccaaaatttgaagagcaaataagaaaaggtggtcctgttacagttactc
atcctgaaattacacgttattttatgttgataccagaagcttgtgaactggtcctacaag
ctggtgctattgcaaaaaattcagaggtctttgtcttagatatggggcaacctgtcaaga
ttattgatcttgctaaacaatttattagactttctggtagaggtgatattgatattaaaa
tagttggtttgcgtccaggagagaaactttacgaagagcttttgatagaggaagatgatg
ttagtaccgactataaagatatttttattggtagaaggacttttacgatattaatactc
taaaccaagatattgaatcgttgatcaaggatgatgttgatcagcttgtgatattaaaga
aaattgttccggaatttgaacatagattgaatgggtagtggttttatgttttatgaggtt
tttaaaagattgcttgatattttactttcttttatggggttgttgttattaagtcctatt
ttcttaattattatttttatgataaagaaagattcaaaaggacctatatttttaaacaa
aagcgctatggtaaagataagcaatttttttacatatataagtttagaactatgtatgtt
gatactccaaaagatatgccaacgcacatgttacaggatccatcgaaatgtataactaag
gttggaggatttttaaggaaatcatctttagatgagttgccacaaattataaatattcta
aaaggtgaaatgagcatcgtgggtccaagaccagcattatggaatcaagatgacttaata
gcacaaagagataagtatggggcaaatgctgtgcctgtgggactgactggctgggcacag
attaatggtagggatgaattaccaatacctgataaagctaaacttgatggtgattatgta
aaaaataaaagtacatggtttgatttaaaatgtatttttttgacagtattttctgttttt
gccaaaaagggcgtcgttgagggtggtactggagctttaggtaacaaagaggatttaaag
tagtatgaaaaaaagaatcttagttacaggtttgagtagctatattggtaactcatttgc
ggctaaatataactcagattttagtatcgataaaatatctttgcgcgatgtttcgtgggc
aaatatagacttaagtggttatgatgctgtattgcatgtcgctggaattgcccatacttc
aaaggatcctaaactaaaagaaaaatactataaaataaatacgcaattaacttatgatct
ggcaaaacaagctaaagatcaaggtgttcgacagtttgtgtttttaagtagtattatagt
ttatggtgatagtgcgccaataggtcaacaaaaagttataactaaatataccgaacctaa
accagatgatttttatggagatagtaagcttcaaactgaaattaagctaaatagcctggc
tagtgatgactttaatattgctataatcagaccaccaatggtatatggagaaggctcaaa
aggcaactatccaaagttggttaaacttgcaaagtatacttttatttttcctaatattaa
taaccaaagaagtgttatatctatagataatttatctaaagagattgcagaaataatttt
gcaaactaaacatggagtttttctacttcaagataatgaatatttttgcacttcacagtt

Fig.6 (Cont I).

tataaaaaactatagaaaagatgttttaggtaagagaacttatctgacaaaaatttttaa
tccaattataagattgcttgctaaaaaagtagattttattaataaagtttttgggaattt
gacttatgagaagtaagttattattcatagctaatgattttgatattgtaatatatcgtt
tcagaagagaagtaatcgagtcttttgctgctaaagagtatgagatagtactagtaacac
catattctaagaaagcagaggtttttttgtaaaagtcttggtgttaagtatataaatgttg
atatagatagacgaggcaaaaatccttttaaggatttgcttcttttatttaactatttca
aaataataaaaaaagaaaaacctgattacatttttagctatacaattaaaccaaatttgt
atgttgggttagtgaatttgtttttttaggaagaagttttatccaaatgtaacaggcttag
gaagtgtttttgctaatcatggtattgttcagaagtttataatatctttatataagttat
catttaaaagcaccacaaaagtattctttcagaatgagcaaaataaaaagttatttatag
ctaagaaaataatcagtggagaaaaatcaatattattaccaggttctggggtaaacttag
atgaaaataaatatgttgactatcctaaagaccaaggaatattaaaattcgtttttcttg
gccgaataatgaaagaaaaggggatttatgaattgttagaagcctttgctatacttgaga
aaaaatataaaaatattagtcttgacatttatggtttttgtgatgaaaataaatctaatt
ttatgggaaaggttaatacgataaaatcagtaaaattttatggttttactgataatacta
aagaaaaaatagctagtgcacatgcagttgttttgccatcttaccatgaaggaatgtcaa
atgtgctgttagaagcagctgcgataggtagacctgtaattgcgtcagatattcctgggt
gtagagaaatttttgatgatggtctctctggcttatcatgtaaccctaatgatgtgagtt
ctttacgtaactcattagagcagtttataaatatgtcgtatactgataaaatagctatga
gctataaagctagagctaagatagaaaaagattttgatagaagtattgttgtcaatgcat
acttacagcaaaattaataataagggtttaaattatgagtttatatgaggatatagtcgc
taaaagagaaaaggtttcattggttggcttgggttatgttggtttaccaatagctattgc
atttgcaaaaaaaatagatgtgttaggatttgatatttgtgaaacaaaagttcaacatta
taaggatggttttgatccaacaaaagaagtaggagatgaggctgtcagaaatacgacaat
gaaatttagttgtgatgaaacaagtcttaaagagtgtaaatttcatattgttgcagttcc
tacaccagttaaagcagataaaactcctgatttgacgccgattattaaggcaagtgagac
ggttggtaggaatcttgtcaaaggcgcttatgttgtgtttgaatcaactgtttatcctgg
tgttacagaagatgtttgcgtaccaatacttgaaaaagagtctggcttgaggtctggtga
agatttcaaagttggttactctcctgagaggataaatcctggtgataaggttcataggtt
agaaacaattatcaaagtagtatctggtatggatgaagagtctttagatactatagcaaa
agtttatgagctagtagtagacgcaggagtttatagagctagtagtataaaagtggctga
agctgctaaggttatagaaaactctcaaagagatgttaatatagcttttgttaatgagtt
atcgataatatttaatcagatgggtattgatactctagaggttttagcagcagctgcaac
taaatggaatttcttaaactttaagcctggtcttgttggtggacattgtattggtgttga
cccatattacctaacgtacaaggcagctgagcttggatatcattctcaggtaatattatc
tggtcgtaggataaatgatagtatgggtaaatttgtagttgagaatttagtcaaaaaact
gatatctgcagatatacctgttaagcgagctagagtagcaattttcggctttacttttaa
agaagactgtcctgacactaggaatactcgagttatagatatggtaaaagagctcaacga
gtatggtatagagccatatattatagatccggtagctgataaagaagaggctaaacatga
gtatggacttgagtttgatgatctaagtaaaatggtcaatctagatgcgatcattattgc
tgttagtcacgaacagtttaaagatataacaaagcaacagtttgataggctatatgcgca
taattctagaaagattatatttgacatcaaaggtagtttagataaatctgagtttgaaaa
agattatatttattggagattgtagtggcttacgataatgttaaatttcctcatggttcg
ttttttttggtgactggaggtgcgggttttattggctctaatttatgtgaagttttactt
agtaagggttatagagttaggtgtttagatgatctctcaaatggtcactatcacaatgtt
gagccgtttttaactaattctaattatgagtttataaaaggtgatattagagatttagat
acttgcatgaaagcttgtgaaggtattgattatgttctacatcaagctgcttggggaagc
gtaccaagaagtattgagatgccattagtgtatgaagatataaatgttaaaggtgcatta
aatatgcttgaagcggctagacaaaataacgttaaaaaatttgtctatgcttctagttca
tcagtatatggtgatgagccaaatttacctaaaaaagaaggtagagaaggaaatgtttta
tcaccctatgcatttacaaagaaagctaatgaagagtgggcgagactatacacaaagtta
tatggtctagatacttatggtctaagatattttaatgttttcggtagaagacaagatcct
aatggtgcgtatgcagcagttatacctaaatttatcaaacagttattaaatgatgaagcg
ccaactataaatggagatggtaaacagtcgagagattttacatatatagagaatgttatt
gaggcaaatcttaaagcatgtttagcagatagtaagtatgccggagagtcttttaatata

Fig.6 (Cont II).

gcttatggaggtagagagtatcttatagatttgtactataatctttgtgatgccttgggt
aaaaaaatagagccaaattttggtccagatagagcgggtgatattaagcatagtaatgct
gatatttcgaaggctaggaatatgctcggatataatccggaatatgattttgaattaggc
ataaagcatgctgttgagtggtatttaattaattaaatggtattttaatcaagtgtacat
aaaaaaagtgtcttttaaaatttatatttatatttactagcttttgtattatttttag
tttagaatttaaatttgctatattgaatattatagtttatcttccggcttgtattttggg
ttttttagctcttaaaaaactatttgtcggaaatattgttaagaaacaattagctttcct
ttttttctttttctttttatcaatgatttatttaataatagtccaaataatcttacttga
tgcagcatcattgtttcctcagtttttatttaacattttgatcgcgataggtttttgtaa
ctttatttttgtttcatatgataataatgaaaattatttttttaatatgtctaaaataat
atttttttgttactttcttacaatctatttttgtatttctttcaaggtattatatattttt
aaatgattggatattctttttttttagtgaaaaaagggaatattgagatttcgaatgttat
tgaatataagttaagagtattcggacttagtaacgctggaggggatggtttaggattttc
aattactataggattatgtttttctatattttatttttatcaaatatattaaaggtaaatc
tatatttaccaaacttatgctgtttgtacctttaattcttattgtgttttctaatatttt
catatctagaacatcactcttaacttcttcacttatattgttaataacaatattttatat
atatattaaaaaagaaaaattactgtttattataatattggcgctattctttttatcaat
atggatattgttcaaattaaatttgaatttgagttgggcttttgaaaatatttactcgta
cattcaatctggcgatttttcacatggaagtctaagtgttttaatcaataaaatgctttt
tgtgccagataacctttttgacttggatatttggttgtgaggatgttagtaatactgatat
tggttatattaaatatttatactattatgggattatatttagtatgtttttttatattct
tattattttcttgtactttgaaatgagaaaatgttttatattttcagagtatcgatcatt
atttctattgttgttaatagtatgtttagtttttcaagcaaaaataatttttttgacagt
aggattatttactaaattaaccattatattatttattttttctcttaaagaaaacagctt
tacaactaggagtgtgatttgaaaaggtttgtacatttaataataaaccttaaccaaggt
ggtgctgaaacaatgctttataaactttgcaaatctatggataagtcaatatatcatatt
acgattatatcacttatgggtaggggagtatttgcaaataagttagaagcttatggtgtt
aaagtttatacattaaatttaaataaatttaatgtactatttgtattgtttaaatatatt
aagattatcagaagaataaagcctgatgttattcatgcttggatgtatcatgcaaatgta
atttctatattatgcaagccttttttatagaaagactaaatatataaatagtataagaatg
ggattggagaattatgatggtcataagaatcttacaaagtttatgataaagttgaatgca
aaattttctaagttctcagatttaacattaaataattcaaagaaatcattagaagatcat
caaaatataggttttaaaaaccaatgctttatagcaaatggttttgataaagatgttttt
aaaccgagcttttaaagtatgaaaaatttcgtttaaataatgatttagatgataatgtt
aaaattataggtatcatagcaagaaatcatgctgataaaaatatttctcgtttcttacaa
atagctaatttattgttaaaaagtaatcctagtttacggtttttaattgctggaagagag
tgttcgaaaatagatataggtagttatctagataacaaaagtaatgtaaataagttttt
gtatttgaatctgtggattctagtgaatacttaccagtattagatttatatttgtctaca
tcaaaagttgaaggttttccaaatatacttgcagaagccatgctatgtgaagttcctatt
gttgcttctaatgttggagattgtaaagatatacttaatggatacggtgaagtttttgag
cttagtcaaggtaataaagaaataatagaaaagattatgaaagttttagaaacaacggta
gtcatgaaaaagcgcatgagagaatatataataaataatttttagtatagaagctattttg
gaaaaacacgaaaaactttatcatgagggcagtgtctaatgtgtggagtagtaggctttt
actcatttaataaagaagaaggttttgactcaataattaatcaatcattgctttctataa
agcatagagggtcggatgatagtgggtattggtgcgacaatcaagttactctggggcata
ctagattatcaatacacgatataactaatgcgggacatcagccaatgttatctaatagcg
gtaatactgctattgtgtttaatggagaaatatataattacttatccataaaaaatcagc
tattaagtgaatattcaaatcttaaatttaaaagtaacagtgatactgaggttttggtca
atgctattgaactttggggtatagataaaactttagaaaaatgcataggaatgtttgctt
ttggagtttacagtagaaaaactagttgcttaatactagctagagatagatttggcgaga
agccattatattttggtatccaaaatggtattttgggttttgcatcagaattgaaggcac
ttaagccattaaaggaatgtggctggaggtttgatatagatagagatgctttagcaacat
atatgaggtatgcttatgtaccaacaccatactctatttataaaaatatatctaaactaa
atgtaggtagttacataaaatttgatgctaaaggtaatagtaaagagtataaatattggg
attctaaaaaagtactagattcagaaaaatataaagattcgtatgatcaagcaatcctag

Fig.6 (Cont III).

atttagaaattaagcttaaaagtacactatcaatacaaatgcagtcagatgttcctctag
gagcatttttatccggaggaattgactcaacaactgtagttgctcttatgcaaagtatgt
ctaaagataagataaacacttttagtataggttttaatcaaaaagaatataatgaagctg
agcatgcaagagcagtagcaaaacatataggtacaaaccacacagatatgtatgttacag
aaagagatgctcttgatgtaataccaaaacttgctggaatatatgacgagccctttgctg
attcatcacaaataccaacgtatcttgtgagtaaaatagctaagtcgaaagtaacagttg
cactatcaggtgacgctggtgatgagctctttggcggttataatagatacttttagcac
caaatattgctaaaaaaatcaaatttgctaagttacttaaatatgcaccagatgcttgga
taaaaaaagctgagatattaaattttggtaagttcgctttattagcagataaactactaa
aactaaaaagagttctcgaaaaagcaaaaacaaataaagagctttatgtactactttgtt
cacaaataaatgatactagctttgtgttaggagcaaaagagtatgatatattaagagata
agaatatttatgatattccacaattatctttccaagagtggatgatgtttgttgattcta
atacatatatgatagatgatatattggttaaggttgatagagcagctatggctaactctc
tagagacaagagtgccatttttagatcataatatttatgaatttgcttattccttaccaa
ttgactataaaatacaacgaggtaacggaaaaagaattttgaaagatttgttatataaat
atgtgccagaaagtttggtcaataggtctaagatggggtttggtattccgcttgctaaat
ggttaagagaagatttacgagagtgggcagataatttactggattatagtaaaatagaca
agcaaggttacttaagtcctgaggtggtgcaaaaatattggcaagagcatttgagtggta
aaagaaattggcaagcaatattatggaatattctaatttttcaggagtggttagataatg
agtaaagtaaatgtaacaaaaccatacttaccagatataaataaatataaaagctatgta
aataaaatatacaaaaatggatggcttactaataatggtccgttagtgcaagagctagaa
aaaagacttgcaaagtatctaggtgttaaaaatatagttttagtatcaaatggtacaatt
gcattagaaatcgcgtatagagcgttaggagtcaaaggaagtgcaattactactccattt
tcatttgttgctactacatcttcattggtttctaacaatgtaaaaccagtgtttgttgat
attgatgagaatactctaagtatagacgtctctaaaattaagtatgctattgaagaggat
acttcagctattgtgccagttcatgtgtttggaaatggttgtgaagttgaaaaaatagac
atgctggctaaaaaacataacttaaaagttatttatgatgcagcacatgcttttgatgtt
aagtataagggtgagagtatattaaactatggtgatatttcgacattaagttttcatgca
acaaagatttttcattctattgaaggaggtgcgcttatcattaatgatgatagtcttgtt
gaaaaagttcgttatttcattaattttggtatagaaagctcagaatcaataccttactta
ggtactaatgctaaaatgaatgaatttgaggcggctatgggactttgtgttctagatgat
attatagaaattaagagcaaaaggaaagttattacagagatatatgaggctgggttagat
ggattggtaaagtttcaagaacagaatcagcattctagtaggaattatagctattttcca
gtaatatttaggactgaggaggaacttctcagagtacagaaagcactaatacaaaatgat
ataatatcgcgtagatatttttatccatcattagatagtcttagttatatagagccaaag
cagtatatgccaatctcaagagatatatctaaaagaatattatgtttgccaatttatgca
gagttagaagacgataaaattaataaaataattaataatatcaaagaggtttcctcatga
aaaaaatatttgttgttacagataatagaactattctaagtgattttaaaaatatcattg
gtagtaaaaatgatgtacaggttgattatttttgtagtttcaagagtcaaacttctttttg
ccaaagaaatatataacagtgagattaagccaatagatatgaaaaaaaatggcaatgatc
ttattggtaagtatgatttaggtttttcttgtcattcgaaacaattatttccagcaaaat
tagttaattcagtattatgtataaatattcatcctggacttaatccatataatagagggt
ggtttccacaggtcttctctattataaataaactacctataggagcaactattcatgtga
tggatgaagagatagatcatggagatataatcattcaggaagaagttgaagttaattctt
tcgaaaactcttttgatgtttatgctaaagttcaaaaaaaagaagttgagttgttcacta
aagtcatagatgatattttgaataataagttcactcgaatcaaacctaactccgaaggca
actataattcaattcatgattataaaaacatgtgtgaaattgatttagataaaatagtaa
caatgcgggaagcaattgactatctaagggctatgacacaccctccatataaaaatagtt
atttcattgatgagcatggaaataaagtatttgttgctcttgaacttgaaaagataagtt
agaaaaatgagccttaaaaaaaatacaatatcaaattatataacacaactatatactagc
ttaattggtattgttatacttccttttgtatttacaacatttaagtcatgatgcatttggt
ctgattggtttttttacagtttttcaaacgtggttacggttgttggatgttggtataaca
ccaactttatcaagagaagtggctcatgttagaggtagtactgatgactatcattactta
cgcaagttggttagatcgttagagctatttttcattattgttggtgttctggtatttatt
gtaattagtacacattcaaggtatatatccacctcttggttacatataggctcgctagat

Fig.6 (Cont IV).

gctgatagtgtaagtgtatgtattgcacttatgggtttaatgtttgcattaagatgggtg
tctgatctatatggtggtggtttgcgtggctttgaaagacaggttctttataataattta
agtatcatacaaacgacactacagtttattggtggattattatttatctgctatgtgtct
actaatattatgtattattttgtatatcagacaataattgcgatactatatctagtatgt
attgcaattgcattttataaaatactaccatcatcatttagcgtgggtttaaggtttgat
tttaaaataattagaaaagtgcttccatttgcactaggcattgcatattctacaacagtt
tggattattgtcactcaatctgataaattagtgttctcacatgtattaccattatctgag
tatggttatttatctttattgatagtgatatctagtgctgttacgatattgtcctctccg
attagcatagctattcagcctagaatgacaatgctattagcccaacaaaatgtaaaagga
atggaaagcttatatttaaaatcatccttgatctcaattacttttttatctgctgtagta
acatgtgttttgatgtattctcatcagctgttgcagtcatggacaggaagtatggaaatt
gctaattggggtagtaatatcttaaatatatatgttttatcagcatctattatttgtata
atatcatttcaatatttttacagtatgcttatggtaagttaaagctacataatacatat
aatacaattagtttagtatttttgctcctatagttatatatactgcttataattatgga
gtgtatactacagcactattatggcttggatatgctatagtggggctgataatctggatg
cctattgtacaccatgtatttgctaaaggtatcaataggtatttttttataaatttagca
gttattactatagtatgtttttattatcgttaatatttaagggttggtatatttatcca
agtaaaattgggttggtagaattaatattgattgggtttgcatttttatttatacaaatt
tgtatagagtatgttttgtttcggtacaaggttttgaggtgtatagatgattaaagtttc
agtatgtgtgatgacatacaatcaagaaaagtatattggtcaatgtttagagtctttggt
tactcaagagactgattttgactttgagataatcgttggagatgatttttctacagatgg
tacaagagatgttattcaagagtatcaaaaaaagtatccggatatcataaagccagtttt
tagagataagaatgtgggaattactgaaaatattaaagaaatctattttgttgcaaatgg
tgagtatatagctcatatggatggtgatgattatgcattgcctggtaaacttcaaattca
ggctgatttttggataataatccaagatgtacgggagttttcataatataaatatact
ctatccaaatggtaatatacaacatagtaggtttgcttgttcaaataagagtatattcaa
tttatcagacactttacgcggagttgctgttggtgcaaatagttcaaaaatgttcagaac
atcggttttggatgatttgattttaccggatatagagcttctagattattattttcatgt
tataacagcagaaaaaggttatttaagtttttttaaattctaatgaatcctatagtgtgta
cagaaaaggtattggtatcacatctaagtctaaggaaaaaatctataatacttatgctgg
attatttgaatatttttttggatagatatcctgaagagaaattaaatatttgtatccctgt
tgtgcaaatgataatttcggctattaaagggagatgttttattagtgctattcgtctatt
caaaattttaattagatcaagatgtattccattagtaagttggtttaaatatagatttga
aaaataaatatcatttagaggattatgtgaaatgaagggaataattctagctggtggcag
tggtacaaggctatatccacttaccttgggtgttagcaaacagctgctacctgtttatga
caagccattgttatactatccactatctgtgcttatgcttgcaggtattagggagatatt
aattatctctacagtgcgtgatatctcacttatccaagagcttcttggtgatggttcaca
atttggtatacagttgagttataaaatccagccatcaccagatgggcttgctcaagcatt
tattcttggtgaggagtttttggcgggtgactcagcttgtttgatattaggagataatat
ctactatggtcaaggtatgactacaatgctagagtctgcaagagcacagtgtggaggtcc
agctggtggcgcttgtgtttttggttattatgttaatgatccgcatagatatggtatagt
cgaatttgataagcaaaaaaatgtaatttcggtagaggaaaagccacagaatcctaagtc
acactatgctatcacaggtttatatttttatgataataatgttgttgagtatgctaaaca
agtcaaaccatctgcacgtggtgagctagagattacttcacttaatgagttatatctaaa
agaaaataagctaaatgtcgaactcttagggcgtggctttgcttggcttgatgctggtac
gcatgattcattgctagaggcaggtcaatatgtcgcaactattgagaaaagacaagggct
taaaattgcatgtttggaagaaattgcatggcgtaaaggctttatctcaacacaacaagt
tctagctcaagctgaaaaactttctaagacagagtatggtcagtatctgaagaatttaat
taaggatggtttataaattaatccgtcatacccatgaaggtgggtatctcataaaagttg
gatatgttttggagattccaatctgcgcagtaatgacaggtttggtaatatatagcgatg
ttttacaatgactaaaaatggttttatgtatattcttacaaataaggataatactgttct
gtacatagttgtaacatctaatttgataaaaagaatgtatgagcataaacatagccttgc
agatggttttactaaaaatataatgttaataagttagtttattttgaaatttatgaagat
ataaaagcagcaattctgtgagaaaagcagttgaaaaaatgaaacagatcttggaaagaa
cgaattattaatgagatgaatccgaattggaatgatttatatgaattaatatgtgagtaa

Fig.6 (Cont V).

aacttttgtcttactggtgcagataggtatctctaaatatcagatgtgattgggagatta
ccgcctacgcggtaatgacaagtttatgcggtaatgatagtttagtgagagaatgactag
tcactataggaatgatgatgtaatgaggaatgaaaaaatgaactacaaaccaaaaaatat
cctagtaacaggtgcggcgggatttattggtagtaactatgtgcgtatgatgttatcacg
ctatagtgatatcaaaataatctcgtatgataagcttacttatgcgggtagtttagataa
tctaaaagacttgaataatgaacataaccatacttttataaaaggtgatatttgtgatga
agttttagtatatcaaacactgaaagaatataaaattgatacgatagtacattttgctgc
agaatcgcatgttgataattcaattgctaatccaaaggtatttttagaaacgaatgtgat
aggtacatttacacttttagattgtgctaaaaggtattggttagatgagctaggtttaga
agaaactagttgtaggtttcatcatgtatctactgatgaggtatatggtaccttggcaaa
agatgaaccagcctttactgagattaaggcttatgagccaaattcaccgtattcggcatc
taaggcgggatctgatcatatttctagagcatatcatcatacctataaacttccggtaac
aatttcaaattgttcaaacaactatggaccataccaacatcgagagaaattaatccctgt
agtgataaatagttgtataaactacaagcctattcctgtttacggagatggttcgaatat
tcgagattggctatatgtagaagatcactgcgatgctatccagacaattgttgagaaagg
agtggttggagaggtttataatattggtggtattaatgaagttgataatctaaccttggt
aaaaactatctgtaaactaatggatgaatataaaccagaaaatgctccacattctaactt
aatcacatttgtggaagatagaaaaggacatgattggcgttatgctattgataacagcaa
gattcagaatgagttaggatggaagccatcacaagattttgataagatgtttagacaaac
tattgagttttatctatagcttaaatatttatcttatgagtatctctaaaaaatcaattt
aatttatttttgtgttaaaaagtagtgttcgcaagaatatagttaatccgaaagatattt
gtagaaaaagatatttgtagaaatgttataatgtctaataaaa

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 03/02338

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/52 A61K48/00 C12N15/79 C12N5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EMBASE, BIOSIS, CHEM ABS Data, EPO-Internal, WPI Data, PAJ, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BURROWS L L ET AL: "MOLECULAR CHARACTERIZATION OF THE PSEUDOMONAS AERUGINOSA SEROTYPE O5 (PAO1) B-BAND LIPOPOLYSACCHARIDE GENE CLUSTER" MOLECULAR MICROBIOLOGY, BLACKWELL SCIENTIFIC, OXFORD, GB, vol. 22, no. 3, 1996, pages 481-495, XP002036538 ISSN: 0950-382X table 1 figure 4 abstract --- -/--	1,3,5,6, 10, 15-18, 26-28

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 October 2003	03/11/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hoff, C

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 03/02338

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 41234 A (UNIV GUELPH ;BURROWS LORI (CA); CHARTER DEBORAH (CA); LAM JOSEPH S) 6 November 1997 (1997-11-06) abstract Claims ---	1,3,5,6, 10, 15-18, 26-28
X	PRIOR ET AL.: "Preliminary analysis and annotation of teh partial genome sequence of Francisella tularensis strain Schu 4" JOURNAL OF APPLIED MICROBIOLOGY, vol. 91, - 2001 pages 614-620, XP002255916 abstract ---	1-3,5-8, 10,11, 15,16
X	DATABASE EMBL 'Online! 18 January 2000 (2000-01-18) COWLEY ET AL.: Database accession no. AF147038 XP002256411 -& COOLEY ET AL.: "Isolation and characterization of F.novivida mutants defective in lipopolysaccharide biosynthesis" FEMS MICROBIOLOGY LETTERS, vol. 182, 2000, pages 63-67, XP002255917 the whole document page 65, right-hand column, paragraph 2 -page 66, left-hand column, paragraph 1 -----	1,3,5-8, 10,11, 15,16

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB03/02338

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9741234 A	06-11-1997	CA 2175435 A1	31-10-1997
		AU 2377497 A	19-11-1997
		WO 9741234 A2	06-11-1997
		EP 0904376 A2	31-03-1999
		US 5994072 A	30-11-1999
		US 2003124634 A1	03-07-2003